A COMPARISON OF BEHAVIOURAL PROCEDURES FOR MEASURING VERNIER ACUITY AND GRATING ACUITY

by © Rebecca Rideout

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Science in Experimental Psychology

Memorial University of Newfoundland

September 2023

St. John's, Newfoundland and Labrador

Abstract

I evaluated a novel 4/4 vision testing procedure recently implemented by our laboratory in a larger study to estimate the maturation ages of three visual functions. This rigorous procedure requires participants to locate a stimulus on four consecutive presentations before progressing to more difficult stimuli. I determined the time-efficiency of the procedure and whether children are affected differently than adults by the procedure, which would suggest that maturation age estimates from the larger study are inaccurate. Fifty-five adults and 52 children were tested on grating acuity and vernier acuity using the Teller acuity cards (TAC) and vernier acuity cards, respectively. I hypothesized that the 4/4 procedure would agree with the widely-used staircase and TAC procedures for both visual functions and across age groups, and that the 4/4 procedure would be more time-efficient than the staircase procedure. The 4/4 procedure showed strong to acceptable agreement with the staircase and TAC procedures for grating acuity, while levels of agreement for vernier acuity were poorer due to the misalignment step sizes. For both visual functions, levels of agreement were stable across age indicating that the measurement of children's acuity was not affected differently than the measurement of adults' acuity when tested with the 4/4 procedure. Although for both visual functions, the 4/4 procedure had longer completion times than the TAC procedure, it was more time-efficient than the staircase procedure. Thus, this procedure is sufficiently accurate and practical to measure visual maturation.

Keywords: grating acuity, vernier acuity, Teller Acuity Cards, Vernier Acuity Cards, 4/4 procedure, staircase procedure, Teller Acuity Card procedure.

General Summary

I evaluated a novel 4/4 vision testing procedure that our laboratory developed to be as accurate as the gold-standard staircase procedure and time-efficient like the recommended Teller Acuity Card (TAC) procedure. A sample of 55 adults and 52 children were tested on grating acuity and vernier acuity with the Teller Acuity Cards and the vernier acuity cards respectively, following the 4/4, staircase, and TAC procedures. Although completion times for the 4/4 procedure were longer than for the TAC procedure, it was much more time-efficient than the staircase procedure. The 4/4 procedure showed strong to acceptable agreement with the other procedures on grating acuity, but poorer agreement on vernier acuity due to the misalignment step sizes. The levels of agreement were similar for children and adults. This suggests that this novel technique can be used to measure vision in both children and adults.

Acknowledgements

I would like to express my gratitude and appreciation to my supervisor, Dr. Jamie Drover for his constant support and encouragement throughout my entire graduate school experience. Dr. Drover has provided me with unwavering guidance and reassurance, pushing me to gain experience and confidence as a developing researcher and trusting me to see this project through. I have learned so much from him and would not have been able to get to this point without his expertise and support.

Additionally, I would like to thank Shelley Cornick, Dr. Drover's PhD student who is the PI on our larger project. Learning of her commitment to this project piqued my interest and desire to do the same and put my all into this work. I would also like to thank Alex McDonah, Dr. Drover's honours student, whose honours thesis was a sub-project of my own. I had the honour of working alongside Alex in the lab for parts of the testing component of both of our theses. This dynamic allowed for me to gain experience in taking a leadership role by showing him the ropes of our testing procedure. Testing with Alex brightened the experience and broke up the monotony of the process. I am thankful for his company, help, and positive attitude throughout it all.

Finally, I am forever thankful for the constant support of my boyfriend, Landon, his family, and my own, whose reassurance and encouragement throughout my entire graduate degree quite literally got me to this point (we can do hard things mom!!!). To my dad, thank you for being the ultimate proof-reader, thesaurus, and supporter for 9 years and counting.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
ANOVA	Analysis of Variance
COR	Coefficient of Repeatability
cpd	Cycles per Degree
FPL	Forced-Choice Preferential Looking
logMAR	The Logarithm of the Minimum Angle of Resolution
PL	Preferential Looking
PREP	Psychology Research Experience Pool
SF	Spatial Frequency
TAC	Teller Acuity Cards (II)
VEP	Visual Evoked Potential

Chapter 1: Introduction and Rationale

1.1 Importance of Measuring Visual Function

Our understanding of the visual system is incomplete. It is not yet clear which optical and neural components of the visual system underlie the different visual functions, i.e., our unique visual abilities (Bennett et al., 2019). Specifically, we do not yet know all of the primary limiting mechanisms of vision. Primary limiting mechanisms refer to fundamental processes/structures within the visual system that impose the strictest limitations on a visual function (Skoczenski & Norcia, 2002). A better understanding of the primary limiting mechanisms would be beneficial in terms of visual pathology as it may lead to improvements in preventing, detecting, and treating vision disorders (Kiorpes, 2016; Milling et al., 2014; Niemeyer & Paradiso, 2017). Furthermore, understanding these mechanisms could influence the direction of future vision research, such as the development of models that simulate human vision (Joulan et al., 2015).

To better understand the visual system, we can measure important visual functions across the lifespan to determine their developmental trajectories. These developmental trajectories can be compared to changes that occur within the visual system. Correspondences between the development of visual functions and components of the visual system may reveal structure-function relationships and thus, primary limiting mechanisms. Furthermore, the developmental trajectories of different visual functions can be compared. If functions follow different trajectories, they are likely mediated by different primary limiting mechanisms. If they follow similar trajectories, they may be limited by the same primary limiting mechanisms. Note that this avenue of research requires the development and utilization of tests and testing procedures that can measure important visual functions accurately and effectively across the lifespan.

1.2 Important Visual Functions

To adequately assess, and to achieve a better understanding of the visual system, attention can be directed to the development of important visual functions, such as spatial vision and spatial localization (Atkinson, 1974). Spatial vision is defined as the ability to detect and distinguish patterns and objects from their background and constituent parts (De Valois & De Valois, 1991). Spatial vision is assessed primarily by measuring visual acuity and contrast sensitivity (Braddick & Atkinson, 2011; De Valois & De Valois, 1991). Visual acuity is defined as sharpness of vision and can be determined by estimating the smallest, high contrast pattern that an individual can either recognize (i.e., recognition acuity) or resolve (i.e., resolution acuity; Braddick & Atkinson, 2011; Dobson & Teller, 1977; Olitsky et al., 2002). Contrast refers to the difference in brightness levels between light and dark components of a pattern (De Valois & De Valois, 1991). Recognition acuity is defined as the smallest object that can be discerned (i.e., recognized) in a standard situation (Pointer, 2008). It is often referred to as optotype acuity as it is typically assessed using charts consisting of optotypes, i.e., numbers, letters, or figures usually presented in rows of progressively decreasing size that must be identified at a specified distance (Chapanis et al., 2006; Pointer, 2008). The first optotype acuity test was developed by Snellen in 1862 (Azzam & Ronquillo, 2022; Laidlaw et al., 2003; Pointer, 2008; Rosser et al., 2001). The Snellen chart contains rows of letters that decrease in size from the top to the bottom. Starting at the top, the participant is required to read the letters in the row. If they correctly read a set number of letters in the row, they progress to the next line (Laidlaw et al., 2003; Pointer, 2008; Rosser et al., 2001). The smallest row identified correctly provides a measure of visual acuity. More recently, logMAR acuity charts have been developed to account for the various limitations of Snellen charts. For instance, in contrast to the Snellen chart, logMAR charts have a regular progression of letter size from line to line. LogMAR is a unit of measure of the size of optotype. As one progresses from one row to the next, the size of the optotypes decreases by .1 logMAR unit. A score of 0 logMAR is equivalent to 20/20 in Snellen notation, or normal vision, while a score above 0 logMAR indicates an acuity score poorer than normal vision (i.e., poorer than 20/20), and a score below 0 logMAR indicates an acuity score better than normal vision (i.e., better than 20/20). This and other changes have improved the accuracy and repeatability of the test (Bailey & Lovie, 1976; Laidlaw et al., 2003).

Resolution acuity is defined as the ability to detect, or resolve, a specific object or pattern (Pointer, 2008). The target need not be recognized, but merely detected. It is assessed typically using square wave gratings and thus, is often referred to as grating acuity. Square wave gratings consist of repeating black and white stripes in which the transition from black to white is abrupt (see Figure 1). To test grating acuity, an individual must detect square wave gratings that vary in spatial frequency. Spatial frequency (SF) is an index of the size of the pattern, i.e., the thickness of the stripes, while taking test distance into account (De Valois & De Valois, 1991). It refers to the number of times a full cycle of a pattern (e.g., one light stripe and one dark stripe within a grating) repeats within one degree of visual space and is measured in cycles per degree (cpd; Dobson & Teller, 1977). Thus, low SFs such as 2 cpd represent thick stripes, and high SFs such as 30 cpd represent thin stripes and is the equivalent of 20/20 acuity (i.e., normal adult acuity in Snellen notation). Grating acuity can be assessed using electrophysiological procedures such as the visual evoked potential (VEP) procedure; however, it is most commonly assessed behaviourally using the Teller Acuity Cards (TAC; see Figure 2 for an example of the Teller Acuity Cards and see below for a full description of both procedures; Dobson & Teller, 1977; Hou et al., 2018).

Figure 1

Square Wave Grating Pattern used for testing Grating Acuity

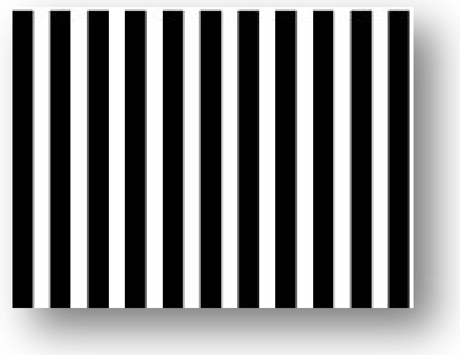
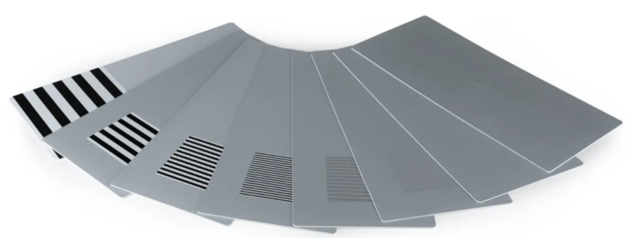


Figure 2

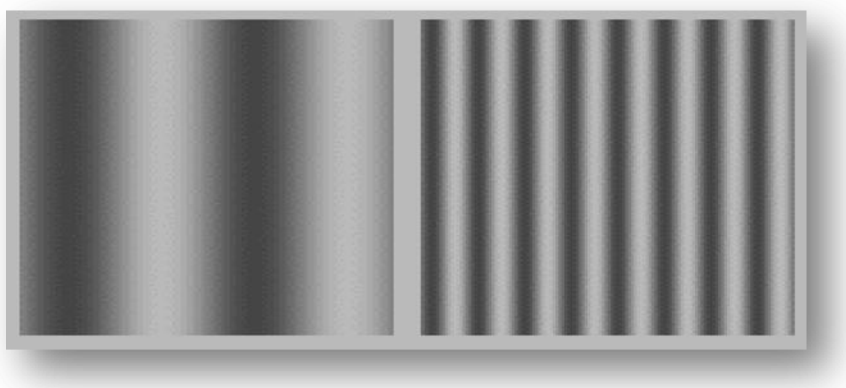
Reference Photo of the Teller Acuity Cards used to test Grating Acuity



Contrast sensitivity is a more comprehensive measure of vision, as it determines one's sensitivity to both size and contrast simultaneously. Contrast sensitivity is measured by determining the contrast threshold that is required to detect sine wave gratings at various SFs (Norcia et al., 1990). Sine wave gratings are those in which the transition from dark to light follows a sinusoidal wave pattern, i.e., there are no abrupt changes in brightness (see Figure 3; Dobson & Teller, 1977). One's contrast sensitivity can be plotted graphically as a contrast sensitivity function, which provides a real-world measure of one's vision as anything under the function can be seen, and anything above it is functionally invisible for that viewer (Norcia et al., 1990). Notably, individuals are most sensitive to intermediate SFs. As with grating acuity, contrast sensitivity can be measured electrophysiologically using VEPs, or behaviourally, using tests such as the contrast sensitivity cards, which were developed based on the TAC (Cornick, 2023).

Figure 3

Sine Wave Grating Pattern used to test Contrast Sensitivity

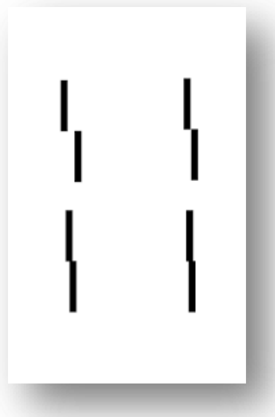


Note: The left image demonstrates low SF, while the right demonstrates high SF.

Spatial localization refers to the process by which the visual system can determine the location of an object in space, by processing and maintaining information regarding its spatial coordinates (De Valois & DeValois, 1991). Spatial localization allows an individual to locate the position of features of a pattern and further differentiate them from the whole (Hu et al., 2021). It can be assessed by measuring vernier acuity, which is the smallest amount of misalignment that can be detected within a stimulus (Hou et al., 2017). More specifically, vernier acuity is measured by determining the smallest offset that can be detected within a line or series of lines (see Figure 4; Hou et al., 2017; Hu et al., 2021).

Figure 4

Sample Stimuli used for testing Vernier Acuity (Bach, 2020).

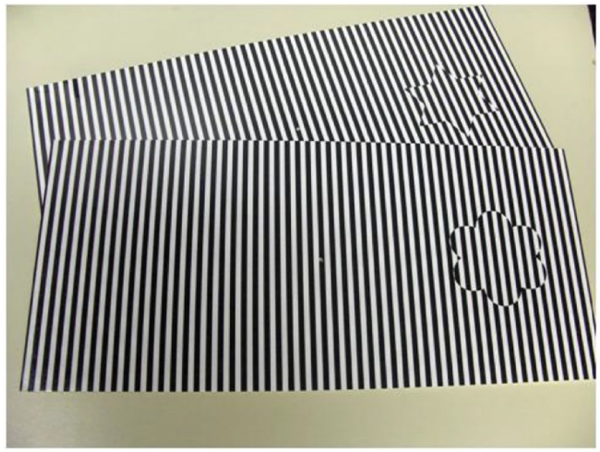


Human vernier acuity is considered a hyperacuity because vernier acuity scores are far greater than optotype and grating acuity scores. Vernier threshold for adult humans has been determined to be as low as 3 to 5 arcseconds, which is equal to 20/1 or 20/2 in Snellen notation (Westheimer, 1987). Additionally, vernier acuity scores exceed the limits imposed by the retina. The reason for this is not completely clear, however, it has been speculated that the combined influence of SF channels and orientation cells beyond the retina in the visual cortex, underlie this hyperacuity status (Levi et al., 2000). Notably, early in life, vernier acuity is poorer than both optotype and grating acuity, however it surpasses both by approximately age 4 (Hu et al., 2021). Vernier acuity can be assessed using behavioural techniques, including the use of vernier acuity cards, a test that is very similar to the TAC the for measurement of grating acuity (see Figure 5;

Drover et al., 2010; Holmes & Archer, 1993). Alternatively, vernier acuity can be assessed electrophysiologically using VEPs (see below for a full description of these procedures).

Figure 5

Reference Photo of the Vernier Acuity Cards (Drover et al., 2010).



1.3 How we Measure Visual Functions

While investigating the development of important visual functions across the lifespan can reveal much about the visual system and its primary limiting mechanisms, to ensure accuracy, it is advantageous to measure each visual function with a single test/procedure. That is, it is beneficial to develop and use a test/measurement technique that is appropriate for adults as well as young children and/or individuals who are not verbal and/or literate (Erhardt et al., 1988). Vision assessment techniques that can be used by individuals of all ages and abilities include electrophysiological procedures and behavioural/psychophysical procedures (Banks & Dannemiller, 1987).

Electrophysiological procedures can determine what one can see by measuring electrical activity in the brain or retina in response to visual stimuli. The measurement of visual evoked potentials (VEPs) is the most commonly used electrophysiological procedure to assess grating acuity, vernier acuity, and contrast sensitivity (Yu et al., 2019). VEP testing is conducted by placing electrodes on the scalp over the visual cortex to measure response amplitudes of cortical neurons when presented with stimuli. Specifically, to measure grating acuity, contrast sensitivity, or vernier acuity, electrical activity is measured in response to visual targets such as square wave gratings, sine wave gratings, or checkerboard patterns that vary in SF/contrast/misalignment, respectively (Dobson & Teller, 1977; Leat et al., 2009). The response amplitudes of electrical potentials are recorded resulting in a regression fit (i.e., data is represented on a straight line that minimizes discrepancies between predicted and actual input; Hou et al., 2018; Leat et al., 2009; Yu et al., 2019). This regression fit is then used to determine the SF, contrast, or misalignment at which the response amplitude is zero, thereby estimating the intercept assuming a linear response. This estimate is subsequently taken as a measurement of visual acuity/contrast sensitivity/vernier acuity, respectively.

Electrophysiological procedures are beneficial as they provide an objective measure of vision, even in nonverbal or illiterate children and adults (Erhardt et al., 1988; Hamilton et al., 2021). However, there are notable limitations of electrophysiological procedures that hinder their usefulness in assessing visual function on a larger scale. The equipment required for these procedures, including VEPs is expensive, costing upwards of 40 thousand USD (Diagnosys, 2022). Measurement of VEPs requires sophisticated equipment and results in relatively complicated output, limiting its use and interpretation of results to expert electrophysiologists (Wright et al., 2012). They also require a considerable amount of time to produce reliable results and have been noted to implement relatively liberal scoring criterion when determining acuity in comparison to techniques with stricter criteria, such as those that rely on behavioural responses/reflexes. This liberality of scoring technique may yield artificially high acuity scores (Dobson & Teller, 1977). Also, VEPs require direct physical contact as skin preparation is required before the attachment of electrodes directly to the scalp to ensure adequate signal-noise ratio, which can cause distress for child participants (Wright et al., 2012).

Behavioural procedures include those techniques that allow the researcher to use their observation of behaviour to determine whether the participants, including nonverbal children, can see a stimulus. Preferential looking (PL) techniques are the most frequently used behavioural techniques to measure visual acuity of young children in laboratory and clinical settings. PL techniques require little comprehension of the test from participants, as all that is required is passive participation and cooperation (Dobson, 1979; Fielder et al., 1992; Mohn & Duin, 1983). These techniques are based on the finding that young children, including infants, will prefer to look at, or fixate on, a pattern as opposed to a blank screen, when presented simultaneously (Fantz, 1958; Fantz 1965). In order to assess grating acuity or contrast sensitivity using PL, the stimulus (i.e.., square wave grating or sine wave grating) is paired with a blank. When assessing vernier acuity using PL, a misaligned stimulus is paired with a non-misaligned stimulus. It is assumed that if the participant can see the stimulus (i.e., square wave grating, sine wave grating, or misaligned stimulus), they will focus their gaze in the direction of the stimulus as opposed to the blank.

Fantz used a PL procedure to measure grating acuity in young children (Fantz, 1958; 1965). In this procedure, the child participant was placed in a test chamber and subsequently exposed to two screens, one of which contained a square wave grating, while the other was a blank grey screen of the same average luminance. An observer was instructed to watch the child's corneal reflections to determine which of the two screens the child fixated. Both the number of times the child fixated each screen, and the duration of each fixation was recorded. The finest stripe width (i.e., highest SF) that produced a significantly greater number of fixations and duration of fixation on the grating as opposed to the plain grey screen, was taken as an individual's acuity estimate.

Teller and colleagues (1979) developed a modified version of Fantz's PL technique, the forced choice preferential looking (FPL) procedure, which uses a two-alternative forced-choice paradigm. While both techniques expose the participant to the same stimuli (i.e., a square wave grating paired with a blank), this new technique differs from the traditional PL technique in that the observer is no longer aware of the position of the grating (e.g., whether it is on the left or on the right). Instead, the observer's role is to determine the position of the grating based on the participant's behaviour, in an effort to reduce observer bias. Looking behaviour is used for infants and non-speaking individuals (i.e., direction of first fixation, duration of fixation, etc.) while more direct behaviour such as pointing or a verbal response is used for testing older individuals who can engage in direct communication with the observer.

Importantly, card-based tests have been developed that can measure visual functions across all ages using these behavioral techniques. The most widely used test of grating acuity is the Teller Acuity Cards (TAC; see Figure 2 above; Drover et al., 2009; McDonald et al., 1985). This test consists of a set of sixteen cards, each containing a square wave grating to the left or right of a central peephole in the card, whereas the remainder of the card is a solid, blank grey and its luminance is the same as the average luminance of the black and white stripes of the grating. Thus, if the SF of the grating is beyond one's threshold, it simply appears as a solid, blank grey 'blending in' with the rest of the card. In all, the set contains square wave gratings across a broad range of SFs (from 0.32 cpd or 1.97 logMAR to 38.0 cpd or -0.10 logMAR). The highest SF grating detected by the participant provides a measure of grating acuity. Though not commercially available, similar card-based tests have been developed for vernier acuity (i.e., see

Figure 5 above, Drover et al., 2010), and for contrast sensitivity (Cornick, 2023). Both card sets are based on the TAC, but instead of a square wave grating, they have either a misaligned target (vernier acuity cards) or sine wave grating (contrast sensitivity cards) on the left or right of the card. Across these sets, the level of misalignment (vernier acuity cards) or contrast (contrast sensitivity cards) varies. These tests can be conducted following the same techniques as for the TAC. The smallest misalignment or lowest level of contrast detected provides a measure of vernier acuity or contrast sensitivity, respectively.

Behavioural procedures are used in tandem with psychophysical procedures, which are methods of estimating thresholds (e.g., smallest, dimmest, etc. stimulus that can be detected) based on behavioural responses (e.g., direction of fixation, tracking, pointing, or verbal response; Bane & Birch, 1992; Banks & Dannemiller, 1987; Cammack et al., 2016; Graham, 1952). Psychophysical procedures use input mapping, which involves varying a stimulus on a particular dimension (e.g., size) in an effort to achieve a desired output or response (Witton et al., 2017). In other words, varying a stimulus across multiple inputs to determine the minimal (e.g., smallest, dimmest, etc.) stimulus that can be detected. Psychophysical techniques are often simplistic, minimizing irrelevant cognitive mechanisms and focusing solely on simple instruction and corresponding behavioural responses (Banks & Dannemiller, 1987; Cammack et al., 2016). These techniques can take on various forms with respect to the method of presentation of the stimuli, as well as the response format that is required. However, they are similar in that they generally involve a large number of trials, with some psychophysical experiments involving hundreds or even thousands of trials per participant (Banks & Dannemiller, 1987; Kuroda & Hasuo, 2014; Schwartz, 2010).

The most common psychophysical techniques include the method of limits (i.e., ascending, and descending series), method of adjustment, method of constant stimuli, and the staircase procedure (Banks & Dannemiller, 1987; Kuroda & Hasuo, 2014). The method of ascending limits presents participants with a stimulus at subthreshold levels to begin, systematically increasing the level of the stimulus with each subsequent presentation until threshold is reached (Schwartz, 2010). The method of descending limits takes the opposite approach, commencing at suprathreshold levels of the stimulus and decreasing the level with each subsequent presentation until threshold is reached**.** The method of adjustment allows for increased participant control whereby participants are given access to a dimmer switch, button, or some other tool allowing them to adjust the level of stimulus presentation (Banks & Dannemiller, 1987; Kuroda & Hasuo, 2014). The method of adjustment generally uses ascending series, in which participants gradually increase the level of stimuli to determine the level at which they can just detect the stimulus. However, the method of adjustment has also been used incorporating both ascending and descending series into the overall experiment and averaging the trials to account for any bias caused by either order of presentation (Kuroda & Hasuo, 2014). Notably, this method cannot be used by young children, as they are not able to accurately control the switch in order to obtain their own threshold independently (Banks & Dannemiller, 1987).

As mentioned above, these methods generally involve a large number of trials with multiple presentations at each stimulus level, where a participant’s threshold is determined by the average of these measures (Kuroda & Hasuo, 2014; Schwartz, 2010). This large number of trials presents limitations that may affect the accuracy of threshold estimation. For instance, there is the possibility of anticipation in ascending series, as the participant may begin to anticipate that they can see the stimulus before they actually can in an effort to be consistent (Cammack et al., 2016;

Schwartz, 2010). Also in some cases, particularly with backlit stimuli, there is the possibility of afterimages in descending series due to the large number of trials over which easily detectable stimuli are presented (Banks & Dannemiller, 1987). Participants may think they can still detect the stimulus, but instead, they see an afterimage of illumination from a previous trial, largely caused by temporary bleaching of photoreceptors on the retina (Li & Sun, 2021; Suzuki & Grabowecky, 2003). The method of constant stimuli is a technique that can account for the limitations of afterimages and anticipation (Banks & Dannemiller, 1987; Schwartz, 2010). Specifically, the experimenter presents stimuli in random order, as opposed to ascending or descending order. Multiple trials are completed, and threshold is estimated as the average of thresholds across individual trials (Banks & Dannemiller, 1987).

However, there remains an inherent problem with the above psychophysical techniques in that they expose participants to many unnecessary stimuli, i.e., those well below or well above threshold, prolonging testing time and leading to boredom or fatigue, and thus affecting performance (Banks & Dannemiller, 1987; Mayer et al., 1982; Schwartz, 2010). There are two solutions to this limitation: the staircase procedure as well as the Teller Acuity Card (TAC) procedure outlined in the TAC II manual.

The staircase procedure is a technique in which the stimulus level presented depends on performance on the previous trial (Atkinson et al., 1982; Witton et al., 2017). It is a combination of ascending and descending limits, generally commencing with a descending series until an error is made. It then shifts to ascending series until the target is detected correctly, after which, descending series presentation then resumes, and so on. These changes in direction are referred to as reversals (Schwartz, 2010). The most common staircase technique follows a two-down one-up procedure, meaning participants must correctly detect a stimulus twice to descend to the next level to a harder to detect stimulus. If they make a single error, they ascend to a higher, easier stimulus level on the next trial and then must correctly detect the stimulus twice to descend again to a harder to detect stimulus (Carkeet et al., 1997). Threshold is estimated as the stimulus level of the last reversal, or the average value of a set number of reversals (e.g., the average of 6 reversals), avoiding any complicated computation (Leek, 2001). This format leads to the stimulus level clustering around the participant's threshold, avoiding irrelevant stimuli, and maintaining accuracy in fewer trials (Banks & Dannemiller, 1987). In addition, the staircase procedure takes significantly less time than the traditional method of adjustment or method of limits (Mayer et al., 1982). Due to this comparatively simplistic and flexible format, the staircase procedure has been adopted as the psychophysical procedure of choice in many laboratories for vision and audition testing and has also been used for testing olfaction and gustation (Leek, 2001; Linschoten et al., 2001). Notably, despite its time-efficiency relative to the above psychophysical techniques, the staircase procedure can still be time consuming in that it can require up to 5 minutes for a child to complete grating acuity testing with the TAC (Chandna et al., 1988). Additionally, the staircase procedure is susceptible to afterimage effects (Banks & Dannemiller, 1987).

The TAC procedure, developed by McDonald and colleagues (1985), is the recommended procedure for use of the Teller Acuity Cards, and can also be used for other card-based tests such as the vernier acuity cards and the contrast sensitivity cards. Unlike traditional psychophysical procedures, the TAC procedure does not require numerous trials and as such, it reduces testing time substantially. Testing, which can be completed in a single trial or few descending series trials, begins with the presentation of an easy to detect stimulus and if it is detected, continues with progressively more difficult to detect stimuli. Following this procedure, each stimulus is presented twice, but if a stimulus is judged by the tester to be difficult to detect, it can be presented up to four times. The most difficult stimulus to detect that is located correctly is taken as an estimate of threshold. The multiple presentations of a single stimulus negates the necessity of multiple trials. Importantly however, the number of presentations is determined by the tester based on the behavior of the participant (e.g., confidence, hesitancy, slow or quick response, etc.). Thus, this procedure is very subjective in that the tester must decide how many times to present each card. This is further complicated by that fact that this decision is based on participants' behavior, as perceived by the individual tester, both of which are subject to individual differences. Similarly, it is considered less rigorous than the above procedures in that if a stimulus is presented only twice, the probability of guessing its location correctly in both presentations is 25%. As with the staircase procedure, the TAC procedure is also subject to the possibility of afterimage effects.

1.4 Purpose of the Present Study

Our research team has been conducting a larger study investigating the development of three visual abilities, grating acuity, vernier acuity, and contrast sensitivity, in participants from age 2 years to adulthood using card-based tests (Cornick, 2023). The purpose of this larger study is to compare the maturation of these visual abilities in an attempt to determine the components of the visual system that underlie them, i.e., the primary limiting mechanisms (Cornick, 2023). Given that multiple visual functions were being measured, we did not use the staircase procedure, as it would be time-consuming and therefore taxing, particularly for young children. Also, because we were attempting to determine the age of maturation of these functions, accurate measurement across all ages was essential. Thus, we did not use the TAC procedure because we did not consider it rigorous enough to be accurate, and as noted above, the probability of guessing the location on two consecutive presentations can be as high as .25. Instead, we developed and used a novel 4/4 procedure in conjunction with descending series in which each stimulus was presented four times and participants were required to detect it on all four presentations in order to move onto increasingly difficult to detect stimuli. We reasoned that this test would be quicker than the staircase procedure while at the same time, more rigorous and objective than the TAC procedure. Specifically, the tester no longer needs to determine which cards are difficult to detect, maintaining a consistent number of presentations throughout. Additionally, the probability of guessing the target location using the novel 4/4 procedure (.06) is substantially less than the probability of guessing the target location using the TAC procedure (.25).

Given that this novel 4/4 procedure requires 4/4 detections per stimulus, one might consider it to be overly conservative, and it is possible that children were affected differently by this conservative criterion than adults. For example, it is conceivable that this procedure was particularly arduous for young children and underestimated their visual abilities relative to adults, thereby impacting the estimates of maturation ages. Specifically, because approximately five cards were used for testing each grating acuity and vernier acuity, and each card was presented 4 times, the 4/4 procedure could have involved up to 40 card presentations. In the present study, I investigated this possibility by testing a sample of adults and a sample of children on grating acuity and vernier acuity using the TAC and vernier acuity cards, respectively, following three different testing procedures: the staircase procedure, the TAC procedure, and the 4/4 procedure. Acuity scores and completion times across the three procedures were compared, the latter to confirm that the 4/4 procedure is more time-efficient than the staircase procedure. More importantly, given the accuracy of the staircase procedure

(Banks & Dannemiller, 1987), I used it as a benchmark and determined the level of agreement between acuity scores obtained using this procedure and the 4/4 procedure. I expected good agreement between the procedures. Furthermore, I expected that the level of agreement between these two procedures would be similar for adults and children. This would suggest that children are not affected differently by the procedure than adults and would support the notion that estimates of maturation ages from the larger study are accurate.

Chapter 2: Method

2.1 Participants

A total of 107 participants, ranging in age from 4-76 years old, participated in the present study. Fifty-five participants comprised the adult sample (mean = 34.8 years, SD = 29.7 years, median = 24 years, IQR = 29 years, range 18-76 years), while 52 participants comprised the child sample (mean = 7.8 years, SD = 2.8 years, median = 8 years, IQR = 5 years, range 4-12 years). See Appendix A for frequency tables of adult and child participants by age. Power analyses were conducted using *G*Power*, in which it was determined that a sample size of 28 per group was sufficient to obtain a medium effect. Participants were recruited from the general population through a recruitment poster shared on the Facebook profile of the primary investigator. In addition, undergraduate psychology students were recruited through Memorial University of Newfoundland's Psychology Research Experience Pool (PREP), a system that encourages research participation from undergraduate students by awarding them with course credit in eligible psychology courses. Participants recruited through PREP were awarded one credit point (1%) towards their final grade in the eligible course of their choosing. There was no incentive provided to participants from the general population. The study was approved by the Interdisciplinary Committee on Ethics in Human Research (See Appendix B).

2.2 Materials and Procedure

Prior to grating acuity and vernier acuity testing, participants underwent a vision screening in which they completed a visual acuity test and autorefraction. Most children and adults completed a standard Early Treatment of Diabetic Retinopathy Study (ETDRS) visual acuity test in which they viewed a chart from a distance of 3 m that contained rows of letters that were progressively smaller from top to bottom. Participants were asked to read the letters beginning at the top. The last row of letters read correctly was taken as a measure of visual acuity. Four-year old and most five-year old children were screened using the Lea Symbols, an alternative to the standard ETDRS test that has the participant identify symbols (i.e., house, heart, circle, square) as opposed to letters (Gräf et al., 2000). The smallest row of symbols identified correctly provided a measure of their visual acuity. Note that visual acuity was assessed in each eye monocularly.

Participants were also screened with the Welch-Allyn Spot Vision Screener. This is a camera-like device that measures binocular refractive error, i.e., nearsightedness, farsightedness, and astigmatism. It is held approximately 1 m away from the participant, and a harmless infrared light is shone into their eyes. The beam of light returns to the device, which measures the extent to which it is out of focus, thereby measuring refractive error. This device also assesses gaze deviation and pupil size. Pass/fail screening decisions were based on criteria similar to those used by Cornick (2023) and other screening studies (see Appendix C; Cotter et al., 2015; Rowatt et al., 2007; Donahue et al., 2013; Silbert & Matta, 2014; Silverstein et al., 2009; Schmidt et al., 2004). Those who did not pass the screening (n = 2) were not included in the present study as it was meant to validate the use of the 4/4 procedure used in the previous study, which investigated the normal development of grating acuity, contrast sensitivity, and vernier acuity (Cornick, 2023). Participants with corrected vision (e.g., contacts, glasses, or laser correction) who passed screening were in included in the present study and were tested with correction.

Upon the completion of screening tests, all participants were tested on grating acuity and vernier acuity. Grating acuity was tested using the Teller Acuity Cards II (TAC), which consists of a set of sixteen grey rectangular 25.5 cm by 55.5 cm cards (see Figure 2 above). Each card contains a square wave grating located to the left or right of the center of the card, and the average luminance of the grating matches the background of the card. The objective of the task is to identify which side of the card contains the grating. The SF of the gratings vary across cards and ranges from .32 cpd (1.97 logMAR) to 38.0 cpd (-.10 logMAR) at the test distance of 55 cm. Only four cards, ranging from 13.0 cpd (.36 logMAR) to 38.0 cpd (-.10 logMAR) were used in the present study. Vernier acuity was tested using the vernier acuity cards (see Figure 5 above; Drover et al., 2010). These cards were developed by the Retina Foundation of the Southwest and are not commercially available. This card set consists of 12 25.5 cm by 55.5 cm cards, each covered with a high contrast low SF grating (0.90 cpd or 1.5 logMAR). There is a misalignment in the grating to the left or right of the center of the cards that forms a symmetrical familiar pattern, i.e., a star or flower. Again, the objective of testing using these cards is to identify which side of the card contains the misalignment. Across the set of cards, the level of misalignment varies from 29.8 arc min (1.47 logMAR) to 0.43 arc min (-0.37 logMAR). However, as with the TAC, only four cards 1.7 arc min (0.23 logMAR) to 0.43 arc min (-0.37 logMAR) were used in the present study. This was done in an effort to reduce overall testing time given that two visual functions were assessed using a total of six procedures.

Both grating acuity and vernier acuity testing were conducted using three separate techniques, the staircase procedure, the TAC procedure outlined in the TAC II Manual, and the novel 4/4 descending series procedure. The staircase procedure was used as a benchmark in the present study due to its rigor and accuracy. In the staircase procedure, participants were presented twice with a card containing the easiest to detect target. The participant was instructed to point to the grating. The card was rotated between presentations to potentially change the location of the target. If the participant detected the target on both occasions, they were presented with the next card in the set, which contained a slightly more difficult to detect target. Once again, the target was presented twice and had to be detected on both occasions. Testing continued with cards containing progressively more difficult to detect targets until the participant failed to detect the target on either the first or second presentation. Then cards with progressively easier to detect targets were presented twice each. Testing continued in order of decreasing difficulty until the target was detected twice. Once this occured, cards with progressively more difficult to detect targets were presented. Once again, this change in directions is a reversal. Testing continued in this fashion until eight reversals were obtained, and acuity was estimated as the average value (i.e., SF or misalignment) for the latter six reversals.

The TAC procedure was selected for use in this study because it is the recommended procedure for use with the Teller Acuity Cards. Testing followed a descending series, which started with the presentation of the easiest to detect card and progressed to cards that were more difficult to detect until acuity threshold was determined, where the participant received the score of the most difficult card that they were able to detect. All cards were presented at least twice and were rotated between presentations to potentially change the location of the grating/misalignment. The participant was required to point at the target. More than two presentations of the same card occurred if the tester perceived the card to be difficult to detect.

In the 4/4 descending series procedure, the participant was presented with the card containing the easiest to detect target on four separate presentations. The participant was required to point to the card. The card was rotated between presentations to potentially change the location of the grating/misalignment. If the target was detected on all four occasions, the participant was then presented with a card containing a harder to detect target. If the participant made one error, testing was halted. The hardest to detect target that was detected on all four occasions was taken as an estimate of acuity.

Grating acuity and vernier acuity were tested monocularly with the right eye, as participants covered their left eye, either with a plastic occluder or they wore sunglasses with the left lens covered with black tape. This is a standard procedure in vision research. The order of test (i.e., TAC, Vernier Acuity Cards) was counterbalanced, while the order of testing procedure (i.e., staircase, 4/4, TAC) was randomized. Routine breaks were not given between each test, though short breaks (approximately 30 seconds to 2 minutes) were sometimes encouraged, specifically if it was evident that a participant was getting bored or distracted or if they were complaining of visual fatigue. Both completion times and acuity scores were recorded for each test. All acuity scores were converted to logMAR units for analyses, as is the standard for vision acuity research. Lower scores in logMAR reflect finer acuity.[1]

2.3 Statistical Analyses

For both grating acuity and vernier acuity, analyses were conducted to determine whether there was an effect of testing procedure on acuity scores and completion times. Data were

[1] "Finer" is used in place of other terms such as "better" or "superior" as these latter term might be misinterpreted to suggest that acuity scores obtained using one procedure are more accurate or valid than those obtained using another procedure.

analyzed for the adult sample and child sample separately. The data were first checked for violations of normality of residuals as well as violations of sphericity. Violations of normality were determined through visual inspection of Q-Q plots, and violations of sphericity were identified using the Mauchly's test. If no assumptions were violated, a repeated measures ANOVA was used to analyze the data, followed by Tukey *post-hoc* comparisons in order to determine where specific differences existed across procedures. Alternatively, if normality of residuals was violated, or if both normality and sphericity were violated, the Friedman test was used, which is the non-parametric equivalent to the repeated measures ANOVA. Pairwise comparisons were subsequently conducted using the Durbin-Conover non-parametric alternative to *post-hoc* analyses. If only the assumption of sphericity was violated, the Greenhouse-Geisser correction was used within the repeated measures ANOVA, and Tukey tests were conducted for pairwise comparisons. All analyses were done using Jamovi. Effect sizes were also calculated for all analyses. When the repeated measures ANOVA was used, effect size was calculated in Jamovi using partial eta squared (η_p^2). When the Friedman test was used, effect size was calculated using Kendall's *W* value. For comparisons conducted using Tukey *post-hoc* comparisons as well as the Durbin-Conover non-parametric alternative, effect size was calculated using Cohen's *d*.

The level of agreement of acuity scores between all pairs of procedures (ie., staircase vs. 4/4, TAC vs. 4/4, staircase vs. TAC) was determined for grating acuity and vernier acuity for each age group separately using Bland-Altman coefficient of repeatability (COR) analyses (Altman & Bland, 1983). This COR is an index of the level of agreement between the procedures and is equal to 1.96 times the standard deviation of the difference scores between the two procedures. Thus, these analyses provide the 95% confidence limits for the differences between scores using the two procedures, that is, they represent the 95% limits of agreement. The COR is a standard means of comparing clinical tests, including tests of vision (Altman & Bland, 1983; Bland & Altman, 1986; Reeves et al., 1991; Rijal et al., 2021). CORs were used instead of correlation coefficients because in some instances, the latter is not necessarily an indicator of agreement between tests, but just reflects the testing of a large sample (Altman & Bland, 1983; Bland & Altman, 1986; Reeves et al., 1991). In addition, the results are meaningful in that CORs are in the units of measurement of the actual tests. The lower the COR, the better the level of agreement. It was expected that while there may be significant differences in acuity scores between testing procedures, they would nevertheless agree. The level of agreement was categorized as strong or acceptable based on previous research. Whereas no other study has compared different testing procedures for the same visual acuity or vernier acuity test, multiple studies have compared different visual acuity tests. The range of CORs reported from these tests is .17 logMAR to .40 logMAR (Bellsmith et al., 2022; Chen et al., 2012; Claessens et al., 2023; Leat et al., 2019; Osbourne et al., 2023; Sumalini et al., 2022). Based on the judgment of Claessens et al. (2023), I deemed a COR of .24 or less to be acceptable. A level of agreement of .20 logMAR or less was considered to be strong. This latter designation is based on research by Sumalini et al. (2022) who reported a COR of .20 logMAR when assessing the test-retest reliability of the TAC, i.e., testing participants twice following the same procedure.

The main purpose of the present thesis was to determine whether children are affected differently by the 4/4 testing procedure than adults, as this would suggest that the maturation estimates from the larger Cornick (2023) study are not accurate. Thus, I tested for this possibility in two ways. First, for both grating acuity and vernier acuity, I compared the levels of agreement as indicated by CORs obtained from children to those of adults for all pairs of testing procedures (i.e., staircase – 4/4, TAC – 4/4, staircase - TAC). There is no way to determine the similarity of CORs statistically. Thus, I designated a difference of less than .05 logMAR as indicating good agreement between CORs. This represents a difference of less than half of a line on a visual acuity chart and if two visual acuity scores with such a difference were obtained, they would be rounded to the same line. In addition, we visually compared the Bland-Altman plots of children to those of adults. Bland-Altman plots are illustrations of the level of agreement between the tests. Given the accuracy of the staircase procedure, it was used as a benchmark for these comparisons and as such, I was concerned primarily with comparing the level of agreement of the staircase and 4/4 procedures in adults to that of the children. If the CORs and Bland-Altman plots showing the levels of agreement between the staircase and 4/4 procedures were good overall and similar for children and adults, it would suggest that children were not affected differently than adults by the novel procedure.

Second, I compared the difference scores for all pairs of testing procedures (i.e., staircase – 4/4, TAC – 4/4, staircase - TAC) for children to those of adults. Shapiro-Wilk normality tests indicated that the difference scores for all pairs of procedures violated normality, therefore, these comparisons were conducted using Kruskal-Wallis one-way ANOVAs. Effect size was calculated using epsilon-squared (ε^2). Again, the staircase procedure was used as a benchmark, and my primary concern was whether the staircase – 4/4 difference score for children was different than that of adults. If this was the case, it would suggest that children are affected differently by the 4/4 testing procedure than adults.

Secondary analyses were also conducted to compare adults' CORs, Bland-Altman plots, and difference scores for all combinations of procedures to those from a subsample of younger children. In the larger study, Cornick (2023) estimated the age of maturity of grating acuity and vernier acuity to be approximately 8 years of age. To ensure confidence in these estimates, it is important to demonstrate that children, specifically those who are not yet mature on these visual functions, demonstrate similar levels of agreement between the staircase and 4/4 procedures, as adults who are visually mature. Thus, I compared a sample of children under 8 years of age (N = 22, mean = 4.95 years, SD = 0.21 years, median = 4.5 years, IQR = 1.75 years, range 4 – 7 years) to the sample of adults (N = 55, mean = 34.8 years, SD = 29.7 years median = 24, IQR = 29 years, range 18-76 years) using the procedures described above. Notably, I did not add the children 8 years and above to the adult sample for these secondary analyses, as evidence suggests that a related visual function, contrast sensitivity, is not yet mature (Cornick, 2023), and the visual system overall, is still immature (Garey & De Courten, 1983; Gomez et al., 2019; Hendrickson et al., 2012). Furthermore, we retained the comparisons between adults and 4- to 12-year-old children in this thesis as it allows us to compare results from a sample of children whose visual system is not mature to an adult sample, and it allows us to test for consistency across age groups.

The comparisons between these younger children and adults were conducted as above by inspection of CORs and Bland-Altman plots for both samples, and analyses of difference scores. Shapiro-Wilk normality tests indicated that the difference scores for all pairs of procedures violated normality, and comparisons were conducted using Kruskal-Wallis One-Way ANOVAs. Effect size was calculated using epsilon-squared (ε^2). Again, the staircase procedure was used as a benchmark, and our primary concern was whether the staircase – 4/4 difference score for these young children was different from that of adults. If this were the case, it would suggest that children are affected differently by the 4/4 testing procedure than adults.

Chapter 3: Results

3.1 Acuity Scores

3.1.1 Grating Acuity Scores

The data for adult grating acuity scores violated the assumption of normality of residuals but did not violate the assumption of sphericity. Significant differences in grating acuity scores were found across procedures, $\chi^2(2) = 50.742$, $p < .001$, $W = .461$. Subsequent pairwise comparisons determined that acuity scores for the staircase procedure were significantly finer than those of the 4/4 procedure, Durbin-Conover = 8.900, $p < .001$, Cohen's $d = .868$, as well as the TAC procedure, Durbin-Conover = 7.598, $p < .001$, Cohen's $d = .713$, both demonstrating a large effect. However, there was no significant difference between the acuity scores for the 4/4 procedure and TAC procedure, Durbin-Conover = 1.302, $p = .196$, Cohen's $d = .123$. See Table 1 for mean adult grating acuity scores.

The data for child grating acuity scores violated the assumption of normality of residuals but did not violate the assumption of sphericity. Significant differences in grating acuity scores were identified across procedures, $\chi^2(2) = 44.675$, $p < .001$, $W = .429$, indicating a large effect. Subsequent pairwise comparisons determined that acuity scores for the staircase procedure were significantly finer than those of the 4/4 procedure, Durbin-Conover = 8.636, $p < .001$, Cohen's $d = 1.121$, as well as the TAC procedure, Durbin-Conover = 5.613, $p < .001$, Cohen's $d = .676$, both of which demonstrate a large effect. The TAC procedure also yielded finer acuity scores than the 4/4 procedure, Durbin-Conover = 3.022, $p = .003$, Cohen's $d = .346$, demonstrating a small to moderate effect. See Table 1 for mean child grating acuity scores.

Table 1

Mean Grating Acuity Scores for Adults and Children in logMAR units

	Staircase	TAC	4/4
Adults	0.04	0.12	0.14
	(0.10)	(0.14)	(0.13)
Children	0.01	0.09	0.13
	(0.10)	(0.13)	(0.12)

Note: Standard deviations are in parentheses.

3.1.2 Vernier Acuity Scores

The data for adult vernier acuity scores violated the assumption of normality of residuals, but did not violate the assumption of sphericity. Significant differences in vernier acuity scores were identified across procedure, $\chi^2(2) = 25.857$, $p < .001$, $W = .235$, demonstrating a moderate effect. The staircase procedure yielded significantly finer acuity scores than the 4/4 procedure, Durbin-Conover = 5.563, $p < .001$, Cohen's $d = .991$, demonstrating a large effect and the TAC procedure, Durbin-Conover = 4.079, $p < .001$, Cohen's $d = .723$, , demonstrating a moderate to large effect. However, there was no significant difference between vernier acuity scores for the 4/4 and TAC procedures, Durbin-Conover = 1.483, $p = .141$, Cohen's $d = .180$. See Table 2 for mean adult vernier acuity scores.

The data for child vernier acuity scores violated the assumption of normality of residuals but did not violate the assumption of sphericity. Significant differences were revealed in vernier acuity scores across procedures, $\chi^2(2) = 23.012$, $p < .001$, $W = .220$, demonstrating a small to moderate effect. The staircase procedure yielded significantly finer acuity scores than the 4/4 procedure, Durbin-Conover = 5.112, $p < .001$, Cohen's $d = .735$, demonstrating a moderate to large effect, and the TAC procedure, Durbin-Conover = 4.017, $p < .001$, Cohen's $d = .597$, demonstrating a moderate effect. There was no significant difference between vernier acuity scores for the 4/4 and TAC procedures, Durbin-Conover = 1.096, $p = .276$, Cohen's $d = .145$. See Table 2 for mean child vernier acuity scores.

Table 2

Mean Vernier Acuity Scores for Adults and Children in logMAR units

	Staircase	**TAC**	**4/4**
Adult	-0.24	-0.13	-0.10
	(0.09)	(0.18)	(0.17)
Children	-0.19	-0.10	-0.08
	(0.13)	(0.17)	(0.18)

Note: Standard deviations are in parentheses.

3.2 Completion Time

3.2.1 Grating Acuity Completion Time

The data for adult grating acuity completion times did not violate the assumptions of normality of residuals or sphericity. A significant difference was revealed between completion times across procedures, $F(2, 108) = 456.611$, $p < .001$, $\eta_p^2 = .894$, indicating a large effect. *Post-hoc* comparisons revealed that the staircase procedure yielded significantly longer completion times than the 4/4 procedure, $t(54) = 19.226$, $p_{tukey} < .001$, Cohen's $d = 3.602$, and the

TAC procedure $t(54) = 29.931$, $p_{\text{tukey}} < .001$, Cohen's $d = 5.051$, both demonstrating a large effect. The 4/4 procedure yielded significantly longer completion times than the TAC procedure, $t(54) = 8.165$, $p_{\text{tukey}} < .001$, Cohen's $d = 1.019$, demonstrating a large effect. See Table 3 for mean completion times for adult grating acuity.

The data for child grating acuity completion times did not violate the assumptions of normality of residuals or sphericity. A significant difference between completion times across procedures was identified, $F(2, 102) = 253.091$, $p < .001$, $\eta_p^2 = .832$, demonstrating a large effect. *Post-hoc* comparisons revealed that the staircase procedure yielded significantly longer completion times than did the 4/4 procedure, $t(51) = 19.165$, $p_{\text{tukey}} < .001$, Cohen's $d = 2.969$, and the TAC procedure $t(51) = 20.664$, $p_{\text{tukey}} < .001$, Cohen's $d = 3.874$, both demonstrating a large effect. The 4/4 procedure yielded significantly longer completion times than the TAC procedure, $t(51) = 3.831$, $p_{\text{tukey}} = .001$, Cohen's $d = .610$, demonstrating a moderate effect. See Table 3 for mean completion times for child grating acuity.

Table 3

Mean Completion Time for Grating Acuity for Adults and Children in Seconds

	Staircase	**TAC**	**4/4**
Adults	233.76	55.82	80.85
	(58.38)	(17.46)	(28.47)
Children	181.52	54.35	70.13
	(47.29)	(20.84)	(30.50)

Note: Standard deviations are in parentheses.

3.2.2 Vernier Acuity Completion Time

The data for adult vernier acuity completion times did not violate the assumption of normality of residuals; however, the assumption of sphericity was violated. A repeated measures ANOVA was conducted using the Greenhouse-Geisser correction, where it was identified that significant differences existed in completion times across procedures, $F(1.644, 88.772) = 303.894$, $p < .001$, $\eta_p^2 = .743$, indicating a large effect. *Post-hoc* comparisons determined that the staircase procedure yielded significantly longer completion times than did the 4/4 procedure, $t(54) = 13.778$, $p_{\text{tukey}} < .001$, Cohen's $d = 2.278$, and the TAC procedure $t(54) = 33.253$, $p_{\text{tukey}} < .001$, Cohen's $d = 5.427$. The 4/4 procedure yielded significantly longer completion times than the TAC procedure, $t(54) = 8.450$, $p_{\text{tukey}} < .001$, Cohen's $d = 1.459$. All of which demonstrated a large effect. See Table 4 for mean completion times for adult vernier acuity.

The data for child vernier acuity did not violate the assumptions of normality of residuals or sphericity. A significant difference in completion time across procedures was identified, $F(2, 102) = 319.954$, $p < .001$, $\eta_p^2 = .708$. *Post-hoc* comparisons revealed that the staircase procedure yielded significantly longer completion times than did the 4/4 procedure, $t(51) = 17.296$, $p < .001$, Cohen's $d = 2.849$, and the TAC procedure, $t(51) = 23.828$, $p < .001$, Cohen's $d = 4.015$, both demonstrating a large effect. The 4/4 procedure also yielded significantly longer completion times than did the TAC procedure, $t(51) = 6.358$, $p < .00$, Cohen's $d = .771$, , demonstrating a moderate to large effect. See Table 4 for mean completion times for child vernier acuity.

Table 4

Mean Completion Time for Vernier Acuity for Adults and Children in Seconds

	Staircase	TAC	4/4
Adults	226.60	57.40	105.02
	(59.34)	(15.30)	(48.92)
Children	181.90	53.23	72.90
	(45.81)	(19.64)	(27.64)

Note: Standard deviations are in parentheses.

3.3 Coefficient of Repeatability and Difference Score Analyses

3.3.1 Coefficient of Repeatability Analyses for Grating Acuity for Adults and Children 4-12

CORs and 95% confidence intervals are presented for adult grating acuity scores in Table 5 below. COR analyses indicated strong agreement between the staircase procedure and 4/4 procedure, between the staircase procedure and the TAC procedure, and between the TAC procedure and the 4/4 procedure, with CORs of .20 logMAR, .20 logMAR and .16 logMAR, respectively. These scores indicate 95% limits of agreement between pairs of procedures of ±2 lines or less on a standard visual acuity chart. The Bland-Atlman plots illustrating the level of agreement between procedures are presented in Figure 6. The figure demonstrates that although scores for the staircase and 4/4 procedures, and for the TAC and staircase procedures are very similar at high levels of acuity, the differences between procedures are slightly larger as average acuity decreases.

Importantly, the CORs for children's grating acuity scores were very similar to those of the adults (see Table 5 for children's CORs and confidence intervals). There was strong agreement between the staircase procedure and 4/4 procedure, between the staircase procedure and the TAC procedure, and between the TAC procedure and the 4/4 procedure, with CORs of .18 logMAR, .20 logMAR and .16 logMAR, respectively (see Table 5 below). As with adults, the children's CORs indicate 95% limits of agreement for the pairs of procedures of within ±2 lines on a visual acuity chart. A comparison between the adults' and children's CORs representing the level of agreement between the benchmark staircase procedure and the 4/4 procedure revealed that they differed by only .02 logMAR, which is less than our criterion of .05 logMAR for good agreement between age groups. The levels of agreement between procedures for children are illustrated in Bland-Altman plots in Figure 6. As with adults, scores obtained using the 4/4 and TAC procedures are very similar to those obtained with the staircase procedure for those with high average acuity, but less similar for those with lower average acuity. Note however, the Bland-Altman plots demonstrate that the patterns of agreement between procedures obtained for children is very similar to those obtained for adults.

Table 5

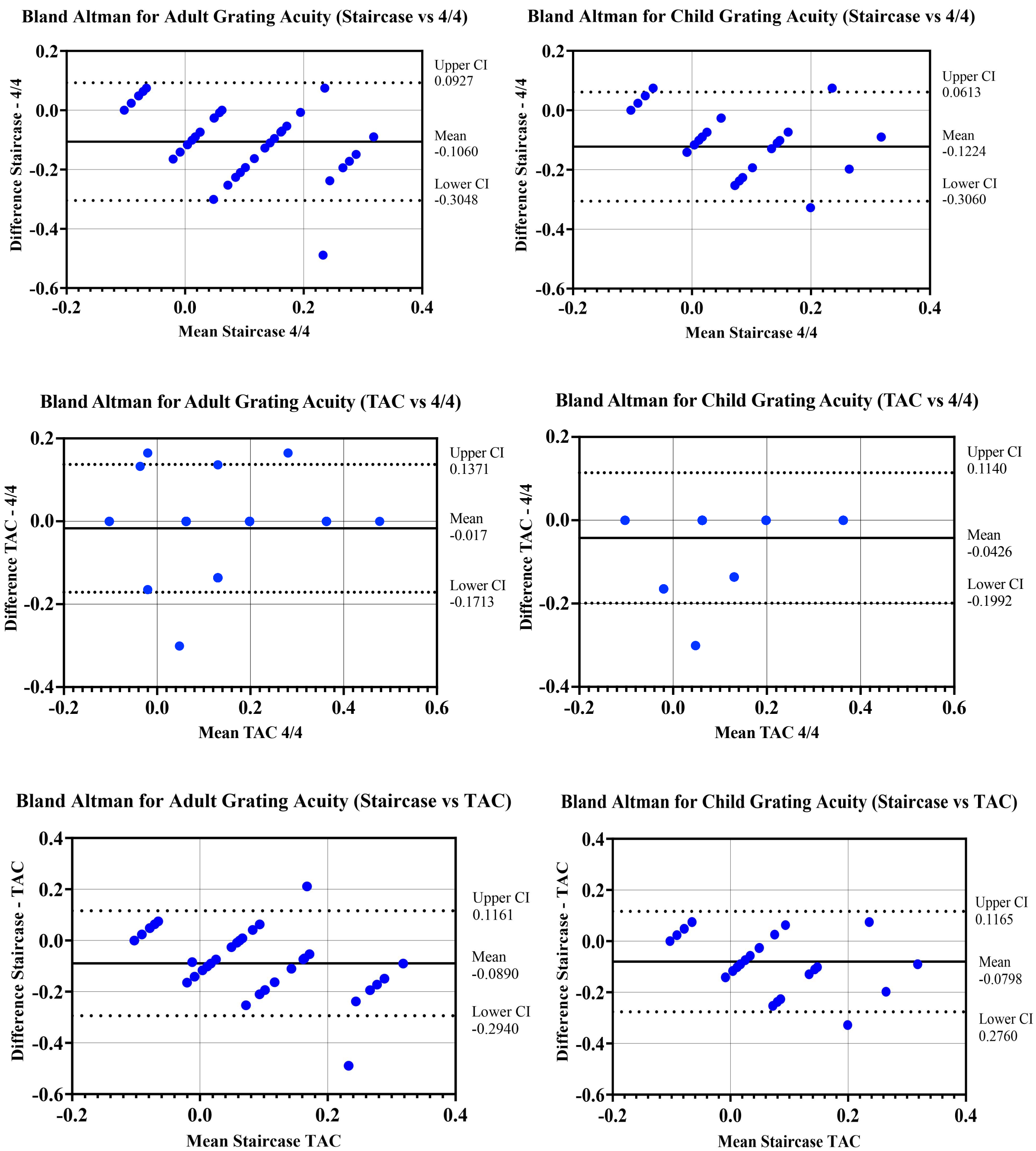
Coefficient of Repeatability for Grating Acuity for Adults and Children 4-12 in logMAR units

	Staircase 4/4	**TAC 4/4**	**Staircase TAC**
Adults	.20	.16	.20
	[.17, .24]	[.13, .19]	[.17, .25]
Children	.18	.16	.20
	[.15, .23]	[.13, .19]	[.16, .24]

Note: 95% Confidence Intervals [Lower, Upper] are in brackets.

Figure 6

Bland Altman Plots for Grating Acuity for Adults and Children 4-12



3.3.2 Difference Scores Analyses for Grating Acuity for Adults and Children 4-12

The difference scores for all pairs of testing procedures are summarized in Table 6 below. For both children and adults, the largest difference scores were for the staircase procedure – 4/4 procedure (adults = -.11 logMAR; children = -.12 logMAR), whereas the smallest difference scores were for the TAC procedure - 4/4 procedure (adults = -.02 logMAR; children = -.04 logMAR). However, Kruskall-Wallis analyses indicated that for all pairs of testing procedures, there were no significant differences between the difference scores of adults and children.

Table 6

Mean Difference Scores for Grating Acuity for Adults and Children 4-12 in logMAR units

	Staircase 4/4	**TAC 4/4**	**TAC Staircase**
Adults	-.11	-.02	-.09
Children	-.12	-.04	-.08

3.3.3 Coefficient of Repeatability Analyses for Grating Acuity for Adults and Children 4-7

In secondary analyses, CORs for grating acuity were also analyzed for a subsample of young children aged 4-7 years, and subsequently compared to CORs from our adult sample (see Table 7 for CORs and confidence intervals). These additional COR analyses indicated strong agreement between the staircase procedure and the 4/4 procedure, and between the TAC and 4/4 procedure at .20 logMAR and .16 logMAR, respectively. The agreement between the staircase procedure and the TAC procedure was slightly lower, but still acceptable at .22 logMAR. These scores indicate 95% limits of agreement for all pairs of procedures of about ±2 lines on a visual acuity chart and importantly, closely resemble the CORs for the adult sample. As with adults and the larger child sample, the CORs between young children and adults showed good agreement as the largest difference between CORs of this sample of children and adults was .02 logMAR. In fact, the CORs of young children and adults was identical for the staircase procedure and the 4/4 procedure, and for the 4/4 procedure and the TAC procedure. A comparison of Bland-Altman plots for adults and young children is provided in Figure 7. As with adults, the young children with high average acuity scored similarly on the 4/4 and staircase procedure, but the difference in acuity between the two procedures was larger for those with lower average acuity. In all, the Bland-Altman plots show similar patterns of agreement for adults and young children.

Table 7

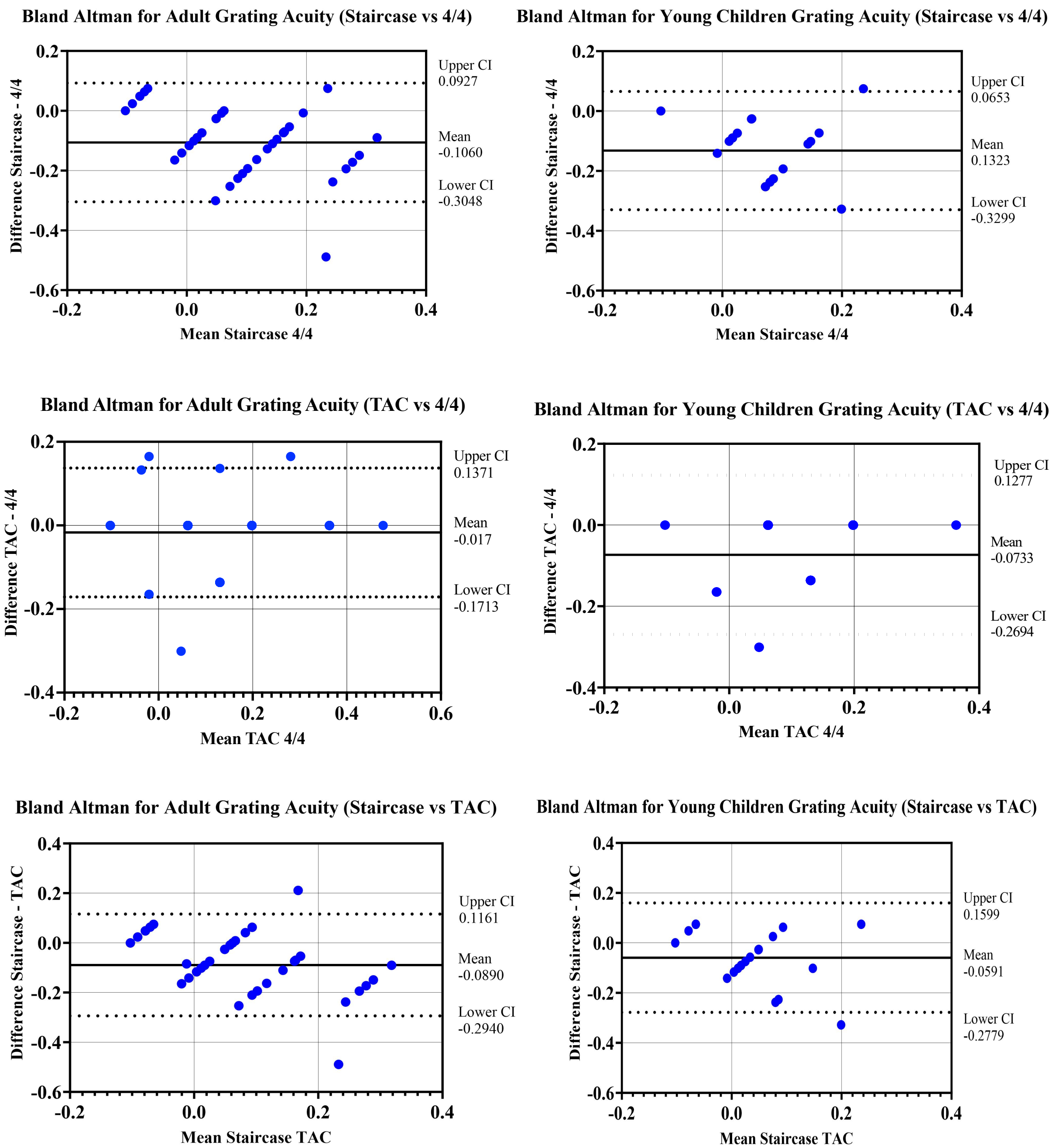
Coefficient of Repeatability for Grating Acuity for Adults and Children 4-7 in logMAR units

	Staircase **4/4**	**TAC** **4/4**	**Staircase** **TAC**
Adults	.20	.16	.20
	[.17, .24]	[.13, .19]	[.17, .25]
Children	.20	.16	.22
	[.15, .23]	[.12, .22]	[.17, .31]

Note: 95% Confidence Intervals [Lower, Upper] are in brackets.

Figure 7

Bland Altman Plots for Grating Acuity for Adults and Children 4-7



3.3.4 Difference Scores Analyses for Grating Acuity for Adults and Children 4-7

The difference scores between testing procedures for adults and young children are provided in Table 8 below. The largest difference scores for both adults and young children were obtained for the staircase procedure – the 4/4 procedure (*M* = -.11 logMAR and *M* = -.13 logMAR respectively). However, a comparison of the difference scores of the young children and adults reveal that they were similar for all testing pairs except for TAC – 4/4. In fact, according to Kruskall-Wallis analysis, the difference between the two age groups for TAC - 4/4 was significant, $X^2 = 5.954$, $p = .015$, $\varepsilon^2 = .078$, indicating a moderate effect. There were no other significant differences in difference scores between adults and young children ($p > .05$).

Table 8

Mean Difference Scores for Grating Acuity for Adults and Children 4-7 in logMAR units

	Staircase 4/4	**TAC 4/4**	**Staircase TAC**
Adults	-.11	-.02	-.09
Children 4-7	-.13	-.07	-.06

3.3.5 Coefficient of Repeatability Analyses for Vernier Acuity for Adults and Children 4-12

According to COR analyses, the level of agreement between procedures for adult vernier acuity scores was not as strong as it was for grating acuity, which is likely an artifact of misalignment step size (see the Discussion section below). This was true for the level of agreement between the staircase procedure and 4/4 procedure, the staircase procedure and the

TAC procedure, and the TAC procedure and the 4/4 procedure, with CORs of .28 logMAR, .29 logMAR and .24 logMAR, respectively (see Table 9 for CORs and confidence intervals). Note that of these CORs, only the level of agreement between the TAC and 4/4 procedure is within the range of what Claessens (2023) considered acceptable. Collectively, the CORs indicate 95% limits of agreement for pairs of procedures of about ±2.5-3 lines on a visual acuity chart. The CORs for adults' vernier acuity are illustrated in the Bland Altman plots provided in Figure 8. Interestingly the plots for the staircase and 4/4 procedure reveal that for those with high average vernier acuity, the 4/4 procedure tends to yield finer scores than the staircase procedure. However, the opposite tends to be true for those with low average acuity. The same general pattern is evident for the level of agreement between the staircase and the TAC procedure.

For child vernier acuity scores, the COR analyses indicated levels of agreement that importantly, were similar to those of the adult sample. CORs for the staircase procedure and 4/4 procedure, the staircase procedure and the TAC procedure, and the TAC procedure and the 4/4 procedure were .30 logMAR, .27 logMAR and .26 logMAR, respectively (see Table 9 for CORs and confidence intervals). None of these levels or agreement are considered acceptable by Claessen's (2023) criterion, which again, is likely an artifact of the misalignment step size. Note that the COR of this sample representing the level of agreement between the staircase and 4/4 procedures differed from that of adults by only .02 logMAR, well within our range of ± .05 for good agreement between age groups. The Bland-Altman plots illustrating the similarity in levels of agreement for adults and children are provided in Figure 8 below. Note that the Bland-Altman plots of children are quite similar to those of adults.

Table 9

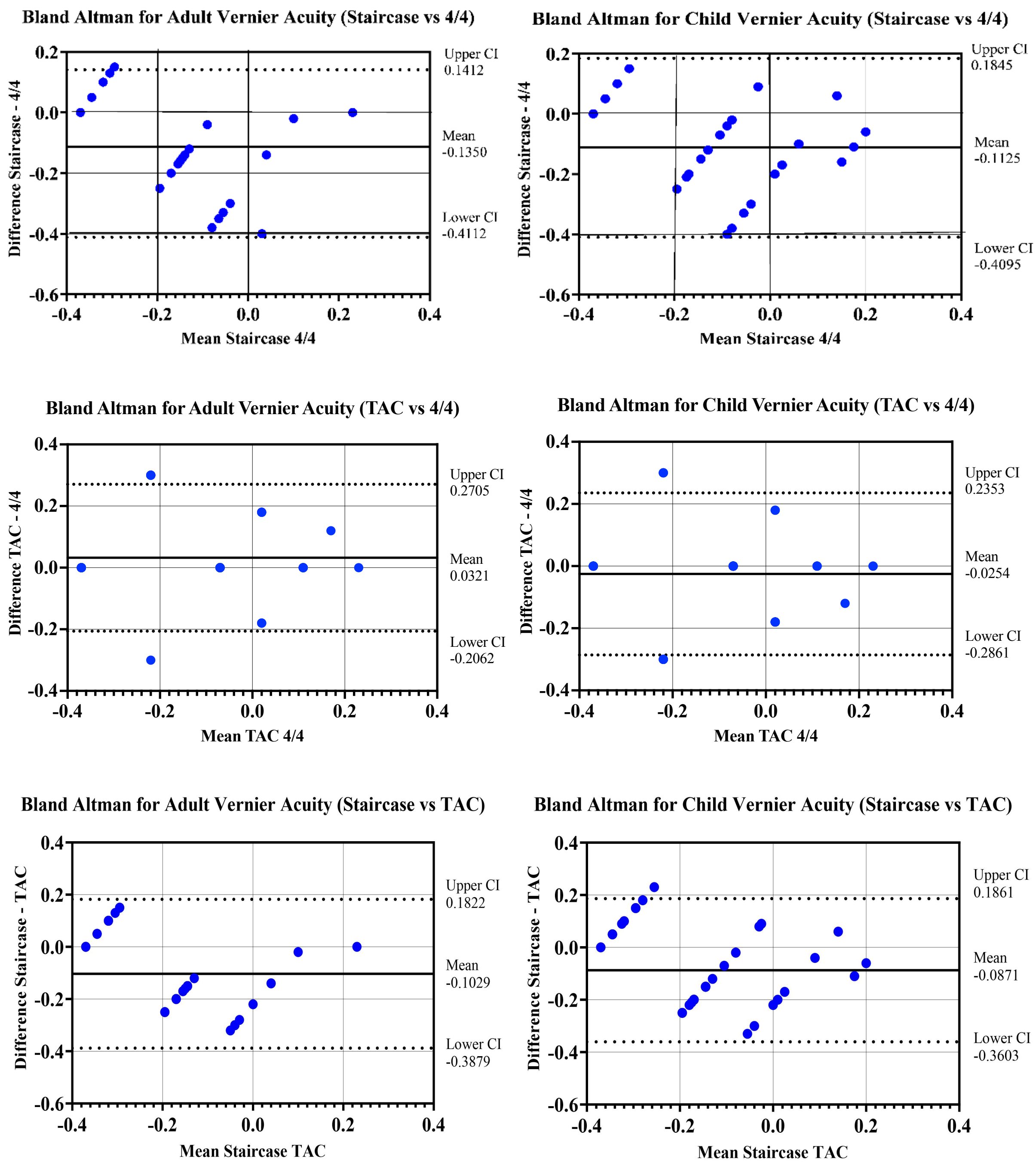
Coefficient of Repeatability for Vernier Acuity for Adults and Children 4-12 in logMAR units

	Staircase 4/4	**TAC 4/4**	**Staircase TAC**
Adults	.28	.24	.29
	[.23, .34]	[.20, .30]	[.24, .35]
Children	.30	.26	.27
	[.25, .37]	[.22, .32]	[.23, .34]

Note: 95% Confidence Intervals [Lower, Upper] are in brackets.

Figure 8

Bland Altman Plots for Vernier Acuity for Adults and Children 4-12



3.3.6 Difference Scores Analyses for Vernier Acuity for Adults and Children 4-12

The difference scores for all pairs of testing procedures are provided in Table 10 below. As with grating acuity, the largest difference scores for both children and adults were for the staircase procedure – the 4/4 procedure (adults: M = -.14 logMAR; children: M = -.11 logMAR), whereas the smallest difference scores were for the TAC procedure – the 4/4 procedure (adults: M = -.03 logMAR; children: M = -.03 logMAR). Yet more importantly, according to Kruskall-Wallis analyses, there were no significant differences between children and adults for any pair of testing procedures (all $p > .05$).

Table 10

Mean Difference Scores for Vernier Acuity for Adults and Children 4-12 in logMAR units

	Staircase **4/4**	**TAC** **4/4**	**Staircase** **TAC**
Adults	-.14	-.03	-.10
Children 4-12	-.11	-.03	-.09

3.3.7 Coefficient of Repeatability Analyses for Vernier Acuity for Adults and Children 4-7

In secondary analyses, CORs for vernier acuity were also calculated for our subsample of children aged 4-7 years, and subsequently compared to CORs from our adult sample (see Table 11 below for CORs and confidence intervals). These additional COR analyses for young children indicated good levels of agreement between the staircase procedure and the 4/4 procedure, and between the 4/4 procedure and the TAC procedure at .24 logMAR and .21 logMAR respectively. The level of agreement between the staircase procedure and the TAC was strong at .19 logMAR, which was better than the level of agreement between the two procedures for adults (0.29

logMAR). This difference of .10 logMAR between CORs is equivalent to a full line on a visual acuity chart. Despite this difference, the COR of the 4–7-year-old children representing the level of agreement between the staircase and 4/4 procedures differed from that of adults by only .04 logMAR, a difference of less than half a line on a visual acuity chart, thereby representing good agreement between age groups. Figure 9 presents the resulting Bland-Altman plots indicating the levels of agreement for adults and young children. The plots illustrate the mean average vernier acuities for young children are lower than those for adults. Also, the plots show that adults who had high average acuities often had finer scores on the TAC or 4/4 procedure than on the staircase procedure. This was not the case for young children who tended to score better on the staircase procedure than on the 4/4 or TAC procedures and had lower average acuities overall.

Table 11

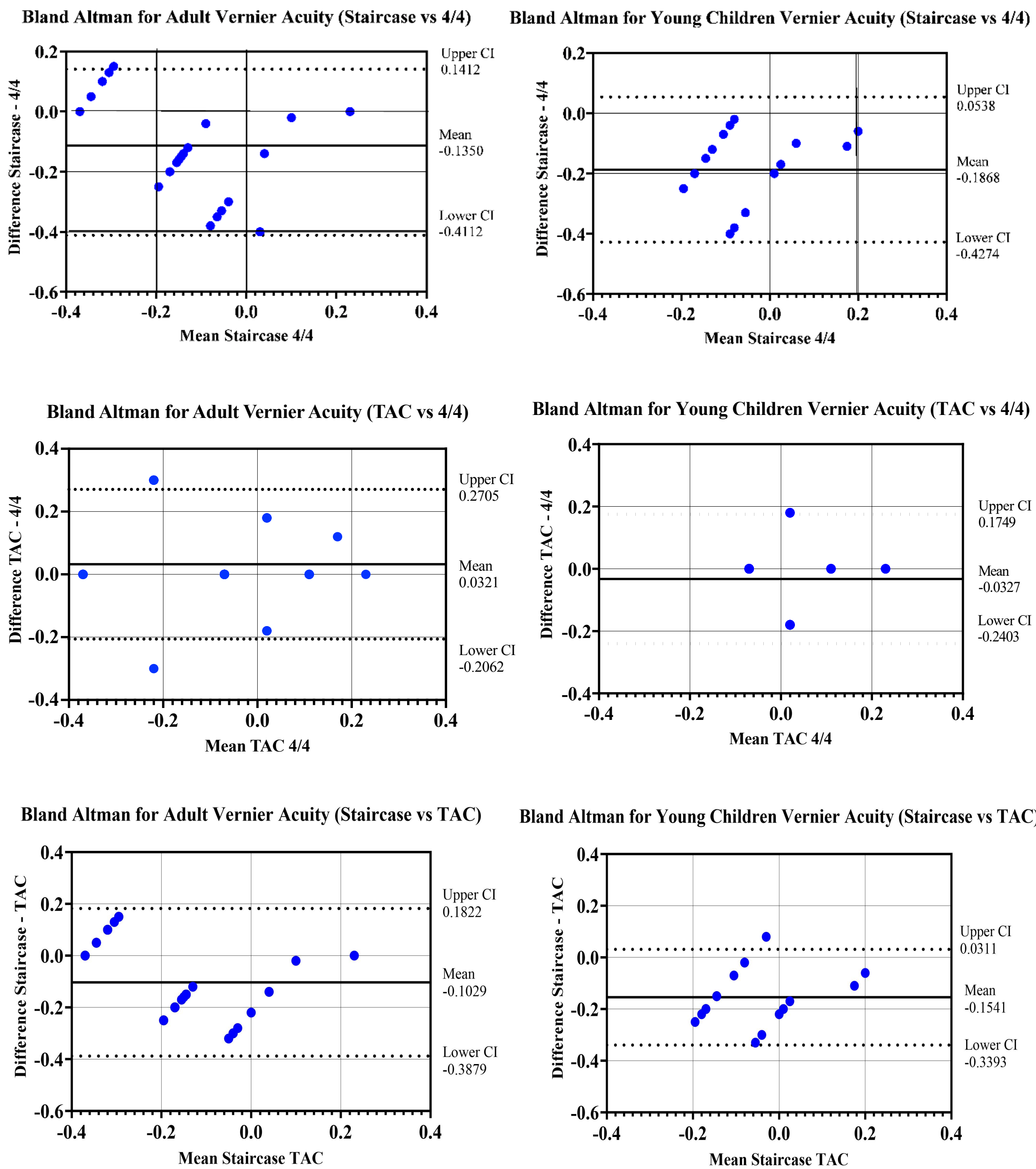
Coefficients of Repeatability for Vernier Acuity for Adults and Children 4-7 in logMAR units

	Staircase 4/4	**TAC 4/4**	**Staircase TAC**
Adults	.28	.24	.29
	[.23, .34]	[.20, .30]	[.24, .35]
Children	.24	.21	.19
	[.19, .34]	[.16, .29]	[.14, .26]

Note: 95% Confidence Intervals [Lower, Upper] are in brackets.

Figure 9

Bland Altman Plots for Vernier Acuity for Adults and Children 4-7



3.3.8 Difference Scores Analyses for Vernier Acuity for Adults and Children 4-7

The difference scores between testing procedures for adults and young children are provided in Table 12 below. Again, the largest difference scores were reported for the staircase procedure - the 4/4 procedure (adults: $M = -.14$ logMAR; children: $M = -.19$ logMAR), while the smallest difference scores were reported for the TAC procedure – the 4/4 procedure (both adults and children: $M = -.03$ logMAR). Once again however, Kruskall-Wallis analyses revealed that there were no differences between adults and the young child group for any procedural pair (all $p > .05$).

Table 12

Mean Difference Scores for Vernier Acuity for Adults and Children 4-7 in logMAR units

	Staircase **4/4**	**TAC** **4/4**	**Staircase** **TAC**
Adults	-.14	-.03	-.10
Children 4-7	-.19	-.03	-.15

Chapter 4: Discussion

In the present study, I evaluated a novel behavioural procedure for testing vision. Specifically, I compared results obtained using the novel 4/4 testing procedure to those obtained using the widely used staircase and TAC procedures. I aimed to determine whether scores yielded by the 4/4 procedure agreed with those yielded by the staircase procedure and the TAC procedure, and whether the novel procedure affected adults and children differently. I also aimed to determine the time-efficiency of the 4/4 procedure. Overall, it was found that while the staircase procedure yielded the finest grating acuity scores, the level of agreement across almost all pairs of procedures was strong. It was also found that this agreement was consistent across age groups. For vernier acuity, the pattern of results was somewhat different. Again, the staircase procedure yielded the finest acuity scores, but the level of agreement across procedures was poor to acceptable. Nevertheless, for almost all pairs of procedures, the level of agreement was consistent across age groups. As hypothesized, for both grating acuity and vernier acuity, the 4/4 procedure was found to be consistently more time-efficient than the staircase procedure, but less time-efficient than the TAC procedure. These results will be discussed in detail below.

4.1 Grating Acuity Scores and Levels of Agreement

In both the adult and child samples, it was found that the staircase procedure yielded significantly finer acuity scores than both the 4/4 procedure and the TAC procedure. The more acute scores yielded by the staircase procedure were expected because participants are able to make detection errors when tested with this procedure, as testing then continues with easier to detect cards, providing the opportunity for additional stimulus presentations. In contrast, once a detection error is made using the TAC or 4/4 procedures, testing is halted, and an acuity estimate is taken based on the most difficult card that was identified correctly. Similarly, the scoring method used for the staircase procedure can lead to finer scores as threshold is estimated as the average stimulus level at which reversals occur. Half of these reversals are cards that were not detected correctly during the presentation, which ultimately leads to finer acuity estimates.

Contrastingly, no significant difference was found between acuity scores for the 4/4 procedure and TAC procedure in the adult sample. However, for the child sample, the TAC procedure was found to yield significantly finer acuity scores than the 4/4 procedure. Notably, the mean difference between these groups was only .04 logMAR, i.e., less than half a line on a visual acuity chart, and the effect size was considered to be small to moderate.

COR analyses demonstrated that the level of agreement between the staircase procedure and 4/4 procedure was strong for all age groups, with values that did not exceed a difference of ±2 lines on a standard visual acuity chart, a point of reference that indicates strong agreement. A very similar pattern was identified for CORs for the staircase procedure and the TAC procedure. Collectively, this indicates that the level of agreement between the 4/4 and staircase procedure is at least as strong as that between the widely used TAC and staircase procedures. The level of agreement between the 4/4 procedure and TAC procedure was also strong for all age groups.

In order to determine whether children were affected differently than adults when tested with the 4/4 procedure, I compared the COR for the 4/4 procedure and the benchmark staircase procedure for children to that of adults. The CORs for the staircase procedure and the 4/4 procedure were similarly strong for adults and children, demonstrating virtually no difference between them and fitting within our criterion for good agreement between age groups. This was true even when comparing the COR of adults to that of younger children (4-7 years) in our secondary analyses. Similarly, an analysis of difference scores demonstrated that staircase – 4/4 difference scores for adults and children were not significantly different. This was also the case when comparing staircase – 4/4 difference scores for adults and younger children. Collectively, these results indicate that the level of agreement between the well-established staircase procedure and the novel 4/4 procedure is strong and consistent across age groups for grating acuity, suggesting that the measurement of children's grating acuity was not affected differently than the measurement of adults' grating acuity when tested with the 4/4 procedure. This supports the notion that the maturation estimates obtained utilizing this procedure in our larger study are likely accurate (Cornick, 2023).

It is noteworthy that the CORs for the staircase and TAC procedures, and for the TAC and 4/4 procedures were also remarkably stable across age. However, there was a significant difference in difference scores for the TAC procedure - 4/4 procedure between the adult group and young children. This contradicts the COR results and the reason for this difference is not clear but, it is worth noting that the 4/4 procedure is considered to be more rigorous than the TAC procedure.

4.2 Vernier Acuity Scores and Levels of Agreement

In both the adult and child samples, it was found that the staircase procedure yielded significantly finer vernier acuity scores than both the 4/4 procedure and the TAC procedure. This pattern of results was again expected due to the nature of the staircase procedure as described in the previous subsection. No significant difference was found between vernier acuity scores for the 4/4 and TAC procedures for the adult or child samples.

COR analyses indicated poor levels of agreement between the staircase procedure and 4/4 procedure, and between the staircase procedure and the TAC procedure for both the adult sample and the child sample, while in the young children, the levels of agreement between those pairs were acceptable and strong, respectively. Nevertheless the level of agreement between the staircase and 4/4 procedures is similar to that between the widely used staircase and TAC procedures. The CORs for the TAC procedure and 4/4 procedure were acceptable for adults and young children, but poor for the larger child sample.

It is clear that the CORs for vernier acuity were poorer than those for grating acuity. In fact, they consistently failed to meet the acceptable criterion of $\leq .24$ logMAR. A likely explanation for this discrepancy can be attributed to the misalignment step size between adjacent cards. Misalignment on the vernier acuity cards is created by pixels, and the smallest level of misalignment cannot be less than a pixel. Thus, in comparison to the rest of the vernier acuity card set and the entire TAC set where the step size between adjacent cards is small (i.e., .12 to .18 logMAR), there is a larger step size from the second most difficult to detect to the most difficult to detect vernier acuity card (.30 logMAR) as the misalignment decreases from two pixels to a single pixel. During vernier acuity testing, it was quite common for a participant, particularly an older child or young or middle-age adult, to detect the most difficult card with one procedure, and the second most difficult card with another, leading to large difference scores, and thereby inflating CORs. The 4/4 and TAC procedures would be most affected by this discrepancy because for these procedures the acuity score aligns directly with the misalignment level of a specific card rather than based on an average score at which reversals occurred as with the staircase procedure. Evidence for this comes from the fact that 4–7-year-old children, who were least likely to detect the highest level of misalignment, were the only age group to have CORs that were deemed acceptable for all procedural pairs.

I should also point out there is no statistical way to determine whether a COR is acceptable, and that any criterion must be chosen at the researcher's discretion. The criterion of ≤ .24 logMAR for acceptable agreement is somewhat arbitrary, and is taken from Claessens et al, (2023) who attempted to validate a web-based visual acuity test by comparing its results to those of a standard Snellen chart. They obtained a COR of .24 logMAR, and considered the test to be valid. It is noteworthy that the range for CORs reported for studies comparing two or more acuity tests is very broad, ranging from .17 to .40 logMAR (Bellsmith et al., 2022; Chen et al., 2012; Claessens et al., 2023; Leat et al., 2019; Osbourne et al., 2023; Sumalini et al., 2022). Thus, the majority of CORs reported here are at the lower end or middle of this range.

To determine whether children were affected differently than adults by the 4/4 procedure, I compared CORs for the 4/4 procedure and the benchmark staircase procedure for children to those of adults. The CORs for the staircase procedure and 4/4 procedure were similar for adults and children as well as adults and young children, both of which demonstrated good agreement. Additionally, an analysis of difference scores for the 4/4 – staircase indicated that the difference scores for adults and children were not significantly different, nor were the difference scores for adults and young children. Collectively, these results indicate good agreement between the well-established staircase procedure and the novel 4/4 procedure that is consistent across age groups. This suggests that the measurement of children's vernier acuity was not affected differently than the measurement of adults' vernier acuity when tested with the 4/4 procedure. Thus, the maturation estimates obtained utilizing this procedure in our larger study are likely accurate.

Notably, the CORs for the staircase and TAC procedures were not stable across all age groups for vernier acuity, despite there being no significant difference in difference scores across age groups. While adults and children demonstrated similar CORs, when comparing adults to young children, the difference in CORs was large (.29 vs .19 logMAR, respectively). Alternatively, CORs for the TAC and 4/4 procedures were remarkably similar across age groups, and no significant difference was found between the difference scores across age group.

4.3 Completion Times

For both grating acuity and vernier acuity completion times, it was found for that for both adults and children, the staircase procedure took the longest time to complete, followed by the 4/4 procedure, with the TAC procedure taking the least amount of time. Notably, the staircase procedure was often multiple minutes longer than the TAC and 4/4 procedures, and while the 4/4 procedure was found to take significantly longer than the TAC procedure, this was often less than 30 seconds in overall difference for grating acuity, and less than a minute in overall difference for vernier acuity. In addition, mean completion times for vernier acuity were longer than for grating acuity. It is possible that this is due to the cards themselves, as participants, particularly adults, often complained of visual fatigue when using the vernier acuity cards, as the high contrast black and white pattern was described as “hard to look at”. When such comments were made, participants were offered to take a break by looking away from the card and opening both of their eyes for a few seconds, prolonging testing time.

These completion times further confirm that while the highly regarded staircase procedure yields the finest acuity scores and is the most rigorous and flexible of the procedures, it was not practical for use in our larger study simply due to the long completion times. In our larger study, child participants were tested using the 4/4 procedure for grating acuity, vernier acuity as well as a third visual function, contrast sensitivity. If this testing routine was followed using the staircase procedure for each participant instead of the 4/4 procedure, testing would end up taking approximately three times longer to complete, resulting in close to an hour of testing per child. This would be difficult for a preschool-aged child to endure, with fatigue and boredom quickly becoming an issue. Similarly, for school-aged children who participated in testing during school hours, it would be unacceptable to disrupt their classroom involvement for an hour.

4.4 Limitations

There were a number of limitations in the present study. First, participants were given the option to participate in the present study in either the MUN vision lab or in their own home or workplace, which caused the room lighting to vary across participants. While the MUN vision lab contains fluorescent lighting that mimics natural light, most participants opted to complete testing at an alternative location where lighting could vary across participants. Notably, only within-subjects' comparisons were made, as each participant completed testing for all three procedures in the same location and under the same lighting conditions, so lighting variance may have increased error (and decreased power), but it should not have biased the results. Additionally, participants were required to complete three procedures for each of the two visual functions, which was particularly taxing for child participants to complete, however signs of boredom and fatigue could be identified in adults as well. As mentioned previously, the differing misalignment step size between adjacent vernier acuity cards was a considerable limitation that affected the levels of agreement between procedures and should be taken into consideration when interpreting the COR analyses results for vernier acuity. Finally, although contrast sensitivity was tested in the larger study (Cornick, 2023), it was not included in the present study. This is because the measurement of this visual function requires estimates of contrast sensitivity at five different SFs. Thus, it is essentially five visual functions in one. Given that the inclusion of contrast sensitivity would require testing at five SFs using three procedures, along with the testing of grating acuity and vernier acuity, testing would have been extremely long and fatiguing. Our laboratory hopes to compare the three testing techniques on contrast sensitivity in the future.

Chapter 5: Conclusion

For both grating acuity and vernier acuity, and across child and adult samples, the staircase procedure consistently yielded the finest acuity scores. This was expected as it is a rigorous procedure that allows participants to make multiple detection errors. Additionally, there is generally no difference in acuity scores across the 4/4 and TAC procedures, with the exception of the significant difference found between these procedures for the 4-12 year-old child sample on grating acuity. The levels of agreement between the staircase procedure and 4/4 procedure were strong for all age groups for grating acuity, but poorer for vernier acuity, likely due to the misalignment step size. Nevertheless, grating acuity and vernier acuity CORs for the staircase procedure and the 4/4 procedure were roughly equal for both adults and children, including younger children. This supports the notion that our estimates of maturation age from the larger study are accurate, since children are not affected differently by the 4/4 procedure than adults. The level of agreement between the staircase procedure and TAC procedure was similar to the level of agreement between the staircase procedure and 4/4 procedure. Although the 4/4 procedure yielded significantly longer completion times than the TAC procedure, completion times were much shorter than for the staircase procedure. Given that it is rigorous, that it does not affect children differently than adults, and its reasonable completion times, the 4/4 procedure can be considered an acceptable procedure for use in our larger study. However, while it can be concluded that the 4/4 procedure is acceptable for use, this does not imply that the TAC procedure is any less suitable for testing grating acuity and vernier acuity. While we originally hypothesized that the 4/4 procedure is a superior option due to the possibility of the TAC procedure being adversely affected by subjectivity and chance agreement, the data ultimately did not suggest that this was true. In fact, the TAC is less time-consuming and might be considered a slightly better option for vision testing when directly comparing the two. Nonetheless, the novel 4/4 procedure remains a viable option for testing grating acuity and vernier acuity in a concise and accurate manner.

References

Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: the analysis of method comparison studies. *Journal of the Royal Statistical Society. Series D (The Statistician), 32*(3), 307-317. https://doi.org/10.2307/2987937

Atkinson, K., Braddick, O., & Braddick, F. (1974). Acuity and contrast sensitivity of infant vision. *Nature, 247*(5440)*,* 403-404. https://doi.org/10.1038/247403a0

Atkinson, J., Braddick, O., & Pimm-Smith, E. (1982). 'Preferential looking' for monocular and binocular acuity testing of infants. *British Journal of Ophthalmology, 66,* 264-268. https://doi.org/10.1136/bjo.66.4.264

Azzam, D., & Ronquillo, Y. (2022). *Snellen Chart.* StatPearls Publishing.

Bach, M. (2020, June 11). *FRACT$_{10}$ Manual*. FrACT$_{10}$ Manual. https://michaelbach.de/fract/manual.html

Bailey, I. L., & Lovie, J. E. (1976). New design principles for visual acuity letter charts. *Optometry and Vision Science, 53*(11), 740-745. https://doi.org/10.1097/00006324-197611000-00006

Bane, M. C., Birch, E. E. (1992). VEP acuity, FPL acuity, and visual behavior of visually impaired children. *Journal of Pediatric Ophthalmology and Strabismus, 29*(4), 202-209.

Banks, M. S. & Dannemiller, J. L. (1987). *Infant Visual Psychophysics*. In Salapatek, P. & Cohen, L. B (Eds.), *Handbook of infant perception: From sensation to perception, volume 1*. Academic Press.

Bellsmith, K. N., Gale, M. J., & Yang, S. (2022). Validation of home visual acuity tests for telehealth in the covid-19 era. *JAMA Ophthalmology, 140*(5), 465-471, https://doi.org/10.1001/jamaophthalmol.2022.0396

Bennett, C. R., Bex, P. J., Bauer, C. M. & Merabet, L. B. (2019). The assessment of visual function and functional vision. *Seminars in Pediatric Neurology, 31,* 30-40. https://doi.org/10.1016/k.spen.2019.05.006

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*(8476), 307-310.

Braddick, O. & Atkinson, J. (2011). Development of human visual function. *Vision Research, 51,* 1588-1609, https://doi.org/10.1016/j.visres.2011.02.018

Cammack, J., Whight, J., Cross, V., Rider, A. T., Webster, A. R. & Stockman, A. (2016). Psychophysical measures of visual function and everyday perceptual experience in a case of congenital stationary night blindness. *Clinical Ophthalmology, 10*, 1593-1606. https://doi.org/10.2147/OPTH.S99593

Carkeet, A., Levi, D. M., & Manny, R. E. (1997). Development of vernier acuity in childhood. *Optometry and Vision Science, 74*(9), 741-750, https://doi.org/10.1097/00006324-199709000-00022

Chandna, A., Pearson, C. M., Doran, R. M. L. (1988). Preferential looking in clinical practice: A year's experience. *Eye, 2*(5)*,* 488-495. https://doi.org/10.1038/eye.1988.98

Chapanis, A., Garner, W. R. & Morgan, C. T. (2006). Applied experimental psychology: Human factors in engineering design. NJ, US: John Wiley & Sons Inc. http://dx.doi.org.qe2a-proxy.mun.ca/10.1037/11152-004

Chen, A., Norazman, F. N. N., & Buari, N. H. (2012). Comparison of visual acuity estimates using three different letter charts under two ambient room illuminations. *Indian Journal of Ophthalmology, 60*(2), 101-104, https://doi.org/10.4103/0301-4738.90489

Claessens, J. L. J., Egmond, J., Wanten, J., Bauer, N., Nuijts, N., & Wisse, R. (2023). The accuracy of a web-based visual acuity self-assessment tool performed independently by eye care patients at home: method comparison study. *JMIR Formative Research, 7*(1), e41045- e41045, https://doi.org/10.2196/41045

Cornick, S. (2023). *It's Complicated: Relationships between structures and functions in visual development.* [Unpublished doctoral dissertation]. Memorial University of Newfoundland and Labrador.

Cotter, S. A., Cyert, L. A., Miller, J. M., & Quinn, G. E. (2015). Vision screening for children 36 to < 72 months: Recommended practices. *Optometry and Vision Science, 92*(1), 6-16. https://doi.org/10.1097/OPX.0000000000000429

De Valois, R. L. & De Valois, K. K. (1991). *Contrast Sensitivity and Acuity.* In De Valois, R. L. & De Valois, K. K. (Eds), *Spatial Vision.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066579.001.0001

Diagnosys. (2022, May 3). *The Leader in Ophthalmic Electrophysiology.* Diagnosys. *https://www.diagnosysllc.com*

Dobson, V. & Teller, D. Y. (1977). Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies. *Vision Research, 18,* 1469-1483.

Dobson, V. (1979). Behavioural tests of visual acuity in infants. *International Opthalamology Clinics, 20,* 233-250. https://doi.org/10.1097/00004397-198002010-00011

Donahue, S. P., Arthur, B., Neely, D. E., Arnold, R. W., Silbert, D., & Ruben, J. B. (2013). Guidelines for automated preschool vision screening: a 10-year, evidence-based update. *Journal of American Association for Pediatric Opthalmology and Strabismus, 17*(1), 4-8, https://doi.org/j.jaapos.2012.09.012

Drover, J. R., Morale, S. E., Wang. Y., Stager, D. R., & Birch, E. E. (2010). Vernier acuity cards: examination of development and screening validity. *Optometry and Vision Science, 87*(11), 806-812. https://doi.org/10.1097/OPX.0b013e3181f6fb5e

Drover, J. R., Wyatt, L. M., Stager, D. M., & Birch, E. E. (2009). The teller acuity cards are effective in detecting amblyopia. *Optometry and Vision Science, 86*(6), 755-759. https://doi.org/10.197/OPX.0b013e3181a523a

Erhardt, R. P., Beatty, P. A. & Hertsgaard, D. M. (1988). A developmental visual assessment for children with multiple handicaps. *Topics in early childhood special education, 7*(4), 84-101. https://doi.org/10.1177/027112148800700409

Fantz, R. L. (1958). Pattern vision in young infants. *The Psychological Record, 8,* 43-47, https://doi.org/10.1007/BF03393306

Fantz, R. L. (1965). Visual perception from birth as shown by pattern selectivity. *Annals of the New York Academy of Sciences, 118*(21), 794-814. https://doi.org/10.1111/j.1749-6632.1965.tb40152.x

Fielder, A. R., Dobson, V., Moseley, M. J., & Mayer, D. L. (1992). Preferential looking – clinical lessons. *Ophthalmic Paediatrics and Genetics, 13*(2), 101-110. https://doi.org/10.3109/13816819209087610

Garey, L. J., de Courten, C. (1983). Structural development of the lateral geniculate nucleus and visual cortex in monkey and man. *Behavioural Brain Research, 10,* 3-13. https://doi.org/10.1016/0166-4328(83)90145-6

Gomez, J., Drain, A., Jeska, B., Natu, V. S., Barnett, M. & Grill-Spector, K. (2019). Development of population receptive fields in the lateral visual stream improves spatial coding amid stable structural-functional coupling. *NeuroImage, 188,* 59-69. https://doi.org/10.1016/j.neuroimage.2018.11.056

Gräf, M. H., Becker, R., & Kaufmann, H. (2000). Lea symbols: visual acuity assessment and detection of amblyopia. *Graefe's archive for clinical and experimental ophthalmology, 238,* 53-58, https://doi.org/10.1007/s004170050009

Graham, C. H. (1952). Behavior and the psychophysical methods: An analysis of some recent experiments. *Psychological Review, 59,* 62-70. https://doi.org/10.1037/h0054020

Hamilton, R., Bach, M., Heinrich, S. P., Hoffmann, M. B., Odom, J. V., McCulloch, D. L. & Thompson, D. A. (2021). VEP estimation of visual acuity: a systematic review. *Documenta Ophthalamogica, 142,* 25-74, https://doi.org/10.1007/s10633-020-09770-3

Hendrickson, A., Possin, D., Vajzovic, L., Toth, C. A. (2012). Histologic development of the human fovea from midgestation to maturity. *American Journal of Ophthalamology, 154*(5), 767-778. https://doi.org/10.1016/j.ajo.2012.05.007

Holmes, J. H. & Archer, S. M. (1993). Vernier acuity cards: a practical method for measuring vernier acuity in infants. *Journal of Pediatric Ophthalmology and Strabismus, 30*(5), 312-314. https://doi.org/10.3928/0191-3913-19930901-10

Hou, C., Kim, Y., & Verghese, P. (2017). Cortical sources of vernier acuity in the human visual system: An EEG-source imaging study. *Journal of Vision, 17*(6), 1-12. https://doi.org/10.1167/17.6.2

Hou, C., Good, W. V. & Norcia, A. M. (2018). Detection of amblyopia using sweep VEP vernier and grating acuity. *Investigative Ophthalmology & Visual Science, 59*(3), 1435-1442. https://doi.org/10.1167/iovs.17-23021

Hu, M. L, Ayton, L. N., & Jolly, J. K. (2021). The clinical use of vernier acuity: resolution of the visual cortex is more than meets the eye. *Frontiers in Neuroscience, 15,* 1-12, https://doi.org/10.3389/fnins.2021.714843

Joulan, K., Brémond, R., & Hautière, N. (2015). Towards an Analytical Age-Dependent Model of Contrast Sensitivity Functions for an Ageing Society. *The Scientific World Journal*, *2015*, 1–11. https://doi.org/10.1155/2015/625034

Kiorpes, L. (2016). The Puzzle of Visual Development: Behavior and Neural Limits. *The Journal of Neuroscience*, *36*(45), 11384–11393. https://doi.org/10.1523/JNEUROSCI.2937-16.2016

Kuroda, T. & Hasuo, E. (2014). The very first step to start psychophysical experiments. *Acoustic Science & Technology, 35,* 1-9. https://doi.org/10.1250/ast.35.1

Laidlaw, D. A. H., Abbott, A., & Rosser, D. A. (2003). Development of a clinically feasible logMAR alternative to the Snellen chart: performance of the "compact reduced logMAR" visual acuity chart in amblyopic children. *British Journal of Ophthalamology, 87,* 1232-1234, https://doi.org/10.1136/bjo.87.10.1232

Leat, S. J., Yadav, N. K., & Irving, E. L. (2009). Development of visual acuity and contrast sensitivity in children. *Journal of Optometry, 2,* 20-26, https://doi.org/10.3921/joptom.2009.19

Leat, S. J., Yakobchuk-Stanger, C., & Irving, E. L. (2019). Differential visual acuity – a new approach to measuring visual acuity. *Journal of Optometry, 13,* 41-40, https://doi.org/10.1016/j.optom.2019.04.002

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics, 63*(8), 1289-1292. https://doi.org/10.3758/BF03194543

Levi, D. M., Klein, S. A., & Carney, T. (2000). Unmasking the mechanisms for vernier acuity : evidence for a template model for vernier acuity. *Vision Research, 40*(8), 951-972. https://doi.org/10.1016/s0042-6989(99)00224-2

Li, H. C. & Sun, P. L. (2021). Visual characteristics of afterimage under dark surround conditions. *Energies, 14(*1404) 1-15. https://doi.org/10.3390/en14051404

Lincschoten, M. R., Harvey Jr., O. L., Eller, M. E., Jafek, B. W. (2001). Fast and accurate measurement of taste and smell thresholds using maximum-likelihood adaptive staircase procedure. *Perception & Psychophysics, 63*(8), 1330-1347, https://doi.org/10.3758/BF03194546

Mayer, D. L., & Dobson, V. (1982). Visual acuity development in infants and young children, as assessed by operant preferential looking. *Vision Research, 22,* 1141-1151. https://doi.org/10.1016/0042-6989(82)90079-7

McDonald, M. A., Dobson, V., Sebris, S. L., Baitch, L., Varner, D. & Teller, D. Y. (1985). The acuity card procedure: a rapid test of infant acuity. *Investigative Ophthalmology & Visual Science, 26,* 1158-1162.

Milling, A., O'Connor, A., & Newsham, D. (2014). The importance of contrast sensitivity testing in children. *British and Irish Orthoptic Journal*, *11*, 9–14.

Mohn, G. & Duin, J. V. (1983). Behavioural and electrophysiological measures of visual functions in children with neurological disorders. *Behavioural Brain Research, 10,* 177-187.

Niemeyer, J. E., & Paradiso, M. A. (2017). Contrast sensitivity, V1 neural activity, and natural vision. *Journal of Neurophysiology*, *117*(2), 492–508. https://doi.org/10.1152/jn.00635.2016

Norcia, A. M. Tyler, C. W., & Hamer, R. D. (1990). Development of contrast sensitivity in the human infant. *Vision Research, 30*(10), 1475-1486, https://doi.org/10.16/0042-6989(90)90028-J

Olitsky, S. E., Nelson, B. A. & Brooks, S. (2002). The sensitive period of visual development in humans. *Journal of Pediatric Ophthalmology and Strabismus, 39*(2), 69-72. https://doi.org/10.3928/0191-3913-20020301-04

Osbourne, D., Steele, A., Evans, M., Ellis, H., Pancholi, R., Harding, T., Dee, J., Leary, R., Bradshaw, J., O'Flynn, E., & Self, J. E. (2023). Children's visual acuity tests without professional supervision: a prospective repeated measures study. *Eye,* 1-6, https://doi.org/10.1038/s41-023-02397-7

Pointer, J. S. (2008). Recognition versus resolution: a comparison of visual acuity results using two alternative test chart optotype. *Journal of Optometry, (2),* 65-70, https://doi.org/10.3921/joptom.2008.65

Reeves, B. C., Wood, J. M., Hill, A. R. (1991). Vistech VCTS 6500 charts- within- and between-session reliability. *Optometry and Vision Science, 68,* 728-737. https://doi.org/10.1097/00006324-199109000-00010

Rijal, S., Cheng, H., & Marsack, J. D. (2021). Comparing the CamBlobs2 contrast sensitivity tests to the near Pelli-Robson contrast sensitivity test in normally sighted young adults. *Ophthalmic and Physiological Optics, 41(5),* 1125-1133. https://doi.org/10.1111/opo.12862

Rosser, D. A., Laidlaw, D. A. H. & Murdoch, I. E. (2001). The development of a "reduced logMAR" visual acuity chart for use in routine clinical practice. *British Journal of Ophthalomology, 85,* 432-436. https://doi.org/10.1136/bjo.85.4.432

Rowatt, A. J., Donahue, S. P., Crosby, C., Hudson, A. C., Simon, S., & Emmons, K. (2007). Field evaluation of the welch allyn suresight vision screener: incorporating the vision in precshoolers study recomenndations. *Journal of American Association for Pediatric Opthalmology and Strabismus, 11*(3), 243-248. https://doi.org/10.1016/j.jaapos.2006.09.008

Schmidt, P., Maguire, M., Dobson, V., Quinn, G., Ciner, E., Cyert, L., Kulp, M. T., Moore, B., Orel-Bixler, D., Redford, M., & Ying, G. S. (2004). Comparison of preschool vision screening tests as administered by licenced eye care professionals in the vision in preschoolers study. *Vision in Preschoolers Study Group. Opthalamology, 111*(4)*,* 637-650. https://doi.org/10.1016/j.ophtha.2004.01.022

Schwartz, S. H. (2010). *Visual Perception: a clinical orientation (4th ed).* McGraw-Hill Medical Pub. Divison.

Silbert, D & Matta, N. S. (2014). Performance of the spot vision screener for the detection of amblyopia risk factors in children. *Journal of American Association for Pediatric Opthalmology and Strabismus, 18*(2), 169-172. https://doi.org/10.1016/j.jaapos.2013.11.019

Silverstein, E., Lorenz, S., Emmons, K., & Donahue, S. P. (2009). Limits on improving the positive predictive value of the welch allyn suresight for preschool vision screening. *Journal of American Association for Pediatric Opthalmology and Strabismus, 13*(1), 45-50, https://doi.org/10.1016/j.jaapos.2008/08.011

Skoczenski, A. M., & Norcia, A. M. (2002). Late Maturation of Visual Hyperacuity. *Psychological Science*, *13*(6), 537–541.

Sumalini, R., Satgunam, P., Subramanian, A., & Conway, C. (2022). Clinical utility of 'peekaboo vision' application for measuring grating acuity in children with down syndrome. *British and Irish Orthopic Journal, 18*(1), 18-26. https://doi.org/10.22599/bioj.264

Suzuki, S. & Grabowecky, M. (2003). Attention during adaptation weakens negative afterimages. *Journal of Experimental Psychology: Human Perception & Performance, 29*(4), 793-807, https://doi.org/10.1037/0096-1523-29.4.793

Teller, D. Y. (1979). The forced-choice preferential looking procedure: a psychophysical technique for use with human infants. *Infant behaviour & development, 2,* 135-153. https://doi.org/10.1016/S0163-6383(79)80016-8

Westheimer, G. (1987). Visual acuity and hyperacuity: resolution, localization form. *American Journal of Optometry & Physiological Optics, 64*(8), 567-574, https://doi.org/10.197/00006324-198708000-00002

Witton, C., Talcott, J. B., Henning, G. B. (2017). Psychophysical measurements in children: challenges, pitfalls, and considerations. *PeerJ,* 1-22. https://doi.org/10.7717/peerj.3231

Wright, K. W., Ning, Y. & Strube, J. (Eds). (2012). *Pediatric Ophthalmology and Strabismus.* Oxford University Press.

Yu, M., Creel, D. & Iannaccone, A. (Ed.) (2019). *Handbook of Clinical Electrophysiology of Vision.* Cham: Springer International Publishing: Imprint: Springer. https://doi.org/10.1007/978-3-030-30417-1

Appendix A

Frequency distribution of child participants by age

Age Range	Frequency
4 - 5	16
6 - 7	6
8 - 9	12
10 - 11	23
12 - 13	5

Frequency distribution of adult participants by age

Age Range	Frequency
18 - 23	20
24 - 29	12
30 – 35	1
36 - 41	3
42 - 47	2
48 - 53	5
54 - 59	9
60 - 65	1
66 - 71	1
72 - 77	1

Appendix B

Most recent Ethics Approval Documentation – ICEHR

MEMORIAL UNIVERSITY

Interdisciplinary Committee on Ethics in Human Research (ICEHR)

St. John's, NL Canada A1C 5S7
Tel: 709 864-2561 icehr@mun.ca
www.mun.ca/research/ethics/humans/icehr

ICEHR Number:	**20231139-SC**
Approval Period:	December 16, 2022 – December 31, 2023
Funding Source:	
Responsible Faculty:	Dr. Jamie Drover Department of Psychology
Title of Project:	*A Comparison of Behavioural Procedures to Measure Grating Acuity and Vernier Acuity*

December 16, 2022

Ms. Rebecca Rideout
Department of Psychology, Faculty of Science
Memorial University

Dear Ms. Rideout:

Thank you for your correspondence addressing the issues raised by the Interdisciplinary Committee on Ethics in Human Research (ICEHR) for the above-named research project. ICEHR has re-examined the proposal with the clarifications and revisions submitted, and is satisfied that the concerns raised by the Committee have been adequately addressed. In accordance with the *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2)*, the project has been granted *full ethics clearance* for **one year**. ICEHR approval applies to the ethical acceptability of the research, as per Article 6.3 of the *TCPS2*. Researchers are responsible for adherence to any other relevant University policies and/or funded or non-funded agreements that may be associated with the project. If funding is obtained subsequent to ethics approval, you must submit a Funding and/or Partner Change Request to ICEHR so that this ethics clearance can be linked to your award.

The *TCPS2* **requires** that you **strictly adhere to the protocol and documents as last reviewed** by ICEHR. If you need to make additions and/or modifications, you must submit an Amendment Request with a description of these changes, for the Committee's review of potential ethical concerns, before they may be implemented. Submit a Personnel Change Form to add or remove project team members and/or research staff. Also, to inform ICEHR of any unanticipated occurrences, an Adverse Event Report must be submitted with an indication of how the unexpected event may affect the continuation of the project.

The *TCPS2* **requires** that you submit an Annual Update to ICEHR before **December 31, 2023**. If you plan to continue the project, you need to request renewal of your ethics clearance and include a brief summary on the progress of your research. When the project no longer involves contact with human participants, is completed and/or terminated, you are required to provide an annual update with a brief final summary and your file will be closed. All post-approval ICEHR event forms noted above must be submitted by selecting the ***Applications: Post-Review*** link on your Researcher Portal homepage. We wish you success with your research.

Yours sincerely,

Kelly Blidook, Ph.D.
Chair, Interdisciplinary Committee on
Ethics in Human Research

KD/bc

cc: Supervisor – Dr. Jamie Drover, Department of Psychology

Appendix C

Pass/Fail Screening Criteria used to determine eligibility for participation

Welch Allyn Vision Screener Criteria

Spherical Equivalent (SE): Acceptable between +/- 1.5D

Sphere (DS): Acceptable between +/- 1.75D

Cylinder (DC): Acceptable between +/- 1.5D

Difference/ Astigmatism (OD, OS): > 1.00D

Screening criteria for use with Early Treatment of Diabetic Retinopathy Study visual acuity chart

An acuity score on either eye > 0.3 (> 20/40)

Acuity scores for each eye that differed by more than 2 lines

Screening criteria for use with Lea Symbols

An acuity score on either eye > 0.3 (> 20/40)

Acuity scores for each eye that differed by more than 2 lines